Tonal alignment in Tokyo Japanese

Takeshi Ishihara

MA, Sophia University, 1999
MSc, University of Edinburgh, 2000

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy

to
School of Philosophy, Psychology and Language Sciences
College of Humanities and Social Science
University of Edinburgh

September 2006

MA, Sophia University, 1999
MSc, University of Edinburgh, 2000

To my grandfather, Shiro Ishihara

Abstract

A large amount of evidence for regularities of tonal alignment in various languages has been accumulated recently. However, there is still much disagreement on the characterisation and modelling of these alignment regularities. This thesis investigates tonal alignment in Tokyo Japanese with two objectives. One is to provide a thorough description of tonal alignment in Tokyo Japanese, including a well-known phenomenon, *ososagari* ('peak delay'); the other is to contribute to the current understanding of tonal alignment, based on empirical data of tonal alignment in Tokyo Japanese.

Three speech production experiments were performed. The first experiment examined the alignment of the F0 targets at the beginning of initial-accented words, varying the syllable/mora structures of the accented syllable. The results showed that both the F0 valley and peak were consistently aligned with specific segmental landmarks, and that the alignment of the F0 peak depended on the syllable/mora structure of the accented syllable. The second experiment explored how the alignment patterns found in the first experiment were influenced in different speaking modes; the speaking modes of interest were fast speech rate, raised voice, and local emphasis. The results showed that the orderly alignment behaviour found in the first experiment remained intact irrespective of different speaking modes, although different kinds of small effects were found. The third experiment compared the F0 peak alignment of unaccented and non-initial-accented words to those of initial-accented words. The results of unaccented words demonstrated consistent alignment of the F0 peak with a specific landmark, which is comparable to those of initial-accented words. On the other hand, the results of non-initial-accented words showed earlier alignment of the F0 peak for the pitch accent than those of initial-accented words.

The results of the current study as a whole demonstrate consistent alignment of the F0 targets with specific places in the prosodic structure in a language-specific way, which

are rather resistant to changes caused by differences of speaking mode. Further durational analyses, together with the alignment data, also suggest that segments and tones are mutually synchronised with each other. These findings provide further evidence that segmental anchoring is a necessary concept in accounting for alignment regularities.

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Takeshi Ishihara

Acknowledgements

First and foremost, I would like to thank my supervisor, Bob Ladd, for his invaluable advice and everlasting encouragement. I am deeply indebted to his clarity of thought and highly perceptive comments. Every time I had a meeting with him, I felt as if I was shown a new path to proceed which I would have never believed it had even existed before. I was truly inspired. I am also very grateful to my second supervisor, Alice Turk, for her scientific rigor and critical remarks. I have always thought that I am very lucky to have had such a wonderful supervisory duo. Needless to say, the responsibility for flaws in this work remains my own.

I am also thankful to my former supervisors and colleagues in Edinburgh and Tokyo: (alphabetically) Art Blockland, Yoshito Hirozane, Keiichi Kojima, Masahiko Komatsu, John Laver, Tsutomu Sugawara, Shin'ichi Tokuma, Michiko Toyama and Etsuhiro Yahagi. I am particularly grateful to Tsutomu Sugawara who gave me access to the facilities at the Phonetics Laboratory of Sophia University, and to Michiko Toyama who helped me with the recordings conducted there, and also sent me various papers to which I otherwise would have had no access.

With regard to my undergraduate years at Kanagawa University in Yokohama, I would like to thank Toshiaki Fukasawa who first inspired me to study phonetics and gave me a word, 'The present does create the future, but the future you wish create the present', which has always given me confidence.

Special thanks to staff at TAAL: Mike Bennett and Alan Whyte, and to friends in Edinburgh and elsewhere: (alphabetically) Maria Luisa Flecha-Garcia, Christine Haunz, Chris Heaton, Mika Ito, Susana M Cortes Pomacondor, Emi Sakamoto, Hajime Yamauchi, Kayoko Yanagisawa and Wing Chiu Ivan Yuen. Especially, I am really thankful

to Kayoko who read several drafts very patiently in a very short period before the submission and gave me helpful comments, and to Ivan who spent endless hours discussing various aspects of my work with me.

Finally, I should like to thank my parents, Kiyoshi and Michiko, my sister, Aya, my grandparents, Shiro and Hiroko, and my old friends back home, for their moral and financial support over the years. Particularly, Shiro's trust and patience have always been of great support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and background

1.1 Association and alignment

In the autosegmental-metrical theory of tone and intonation (Bruce 1977; Pierrehumbert 1980; Ladd 1996, among others), an intonation contour is seen as a sequence of tones which occurs at well-defined locations in the structure. Some of the tones in an utterance are seen to be associated with specific elements of the segmental string such as mora or syllable (and others with the edges of larger domains such as phrase). This relationship is known as *association* and can be schematically represented as:

(1.1)


where 'T' is a tone, and 'TBU' is a tone bearing unit. The elements on these two levels (*tiers*) are linked via an *association line*.[1] Association represents temporal relationship between tones and tone-bearing units, so it can be interpreted that 'T' phonetically occurs during the temporal interval of 'TBU'. However, it has been reported that the phonetic realisation of such association does not simply follow this relationship, and that the temporal synchronisation (*alignment*) of F0 events with segmental events may vary in complicated ways across languages. It is pointed out that there are at least two types of complications to the relationship between association and alignment (Ladd 2003). For

[1]The tone bearing unit may vary between languages; proposed units include syllable, mora, syllable rhyme, syllable nucleus and vowel. In early work of autosegmental phonology, association is attained by the *well-formedness conditions* (Goldsmith 1975), whereas full association between all tones and all TBUs is not always required. It has been argued that tones may float, i.e. stay in the representation without association (*floating tone*), and more recently that TBUs need not to be associated with tones (*phonetic underspecification*) .

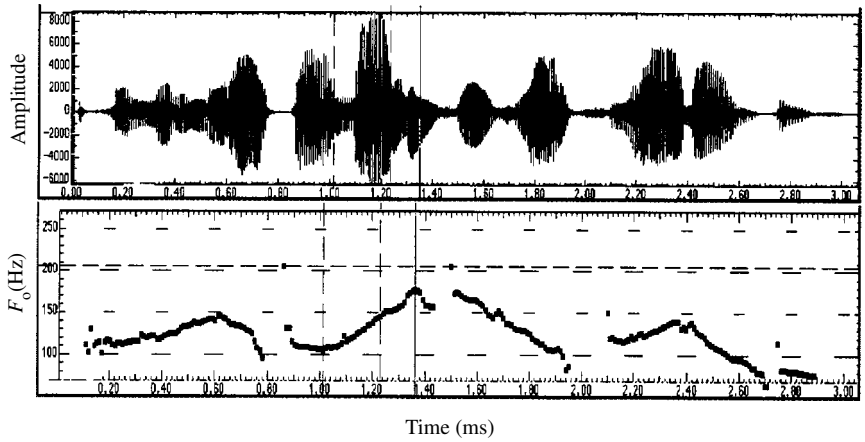


Figure 1.1: An example of peak delay in an utterance, *El murmura autoanálisis raicear antes*, of Mexican Spanish. The first two vertical lines mark the accented syllable (*ná*) and the third line the F0 peak for a pitch accent (H*). Adapted from Prieto *et al.* (1995, p. 430).

one thing, in some languages there appear to be linguistically significant contrasts of alignment. To deal with these alignment differences, bitonal pitch accents are proposed in the autosegmental-metrical theory. A bitonal pitch accent consists of a starred tone which associates with the tone-bearing unit, and an unstarred tone which leads or trails the starred tone. For example, Pierrehumbert (1980) applies L*+H, L+H*, H*+L, H+L* and H*+H (as well as monotonal pitch accents, H* and L*) for the description of American English intonation. An unstarred tone then is considered to be phonetically realised outside the tone bearing unit in languages which have bitonal pitch accents.[2]

For another, there is a phenomenon called *peak delay* in which the F0 peak of a pitch accent may be aligned outside (usually after) the stressed syllable with which it is intuitively associated (an example of peak delay is shown in Figure 1.1). Peak delay has been widely reported across languages,[3] and how it happens seems different from language to language. Much attention has been paid to this phenomenon in the last decade

[2]However, starredness, as well as secondary association, has been recently re-examined and elaborated with further evidence from various languages (Arvaniti *et al.* 2000; Gussenhoven 2000; Grice *et al.* 2000; Atterer and Ladd 2004; Prieto *et al.* forthcoming).

[3]To name a few, (Mandarin) Chinese (Xu 2001), English (Silverman and Pierrehumbert 1990), Greek (Arvaniti *et al.* 1998), Japanese (Neustupný 1966; Sugito 1982), Mexican Spanish (Prieto *et al.* 1995). Although the term 'peak delay' originally refers to the duration from a certain point in the segmental string (e.g. the onset of the syllable rhyme) to the F0 peak of a pitch accent, it is now common to refer to the phenomenon by this name. See, for example, Silverman and Pierrehumbert (1990) for the former use.

or so, partly because it is regarded as an interesting mismatch between phonetics and phonology.

Among the alignment studies, earlier studies focused more on factors affecting the temporal alignment of the F0 targets, (and some of them attempted to model the extent to which those factors affect the alignment, using a multiple regression model). They showed that the F0 target is affected by various factors, such as prosodic boundaries and stress clash. On the other hand, more recent studies have provided various evidence in different languages that the F0 targets of a pitch accent are consistently aligned with a specific point in the segmental string. The following sections are, as a whole, a review of previous studies on tonal alignment. Section 1.2 discusses factors affecting temporal alignment of tonal targets which early studies more attended to. Section 1.3 discusses cross-linguistic regularities of tonal alignment which recent studies demonstrated.

1.2 Contextual factors affecting alignment

An early study on alignment by Silverman and Pierrehumbert (1990) demonstrated factors affecting the F0 peak alignment of the prenuclear H* accent in English. Building on the data of Steele (1986) about the timing of the F0 peak for the nuclear H* accent, they examined the timing of the F0 peak of the prenuclear H* pitch accent under a variety of prosodic conditions (proximity of the following word boundary and of the following pitch accent) and speech rate changes (fast, normal and slow). They modelled the F0 peak proportion ('peak delay' divided by the rhyme duration),[4] as well as the rhyme duration, using multiple regression. The results showed that the F0 peak occurred past the end of the associated syllable, whereas it was systematically shifted earlier by the right-hand prosodic context. For example, with fewer unstressed syllables between the pitch accents, the longer the vowel became and the shorter the peak delay for a given length of the vowel. That is, the too close proximity for the two pitch accents was moderated both by lengthening the vowel and shifting the F0 peak earlier. The results also showed the effect of the word boundary: the vowel was longer and the F0 peak was earlier when the word boundary followed immediately than when there was an unstressed syllable between the prenuclear pitch accent and the word boundary. As for the effect of the speech rate changes, the absolute 'peak delay' (duration from the syllable rhyme onset to the F0 peak) changed—increased and decreased—in correlation to the increase and decrease in rhyme duration induced by the changes in speech rate: the 'peak delay'

[4] 'peak delay' is the duration from the beginning of the syllable rhyme to the F0 peak in Silverman and Pierrehumbert (1990).

was longer when the rhyme duration was longer, and the 'peak delay' was shorter when the rhyme duration was shorter. On the other hand, in terms of the F0 peak proportion, which they used to express the F0 peak alignment, the F0 peak occurred later in fast speech and earlier in slow speech. Overall, the multiple regression models demonstrated that, while the predictors such as word boundary, stress clash, and speech rate contributed to both the peak proportion and the rhyme duration, the predictors' contributions to them were not in the same direction in the two models, and the degree of their contributions were different between them—the factors affected both the peak location and the rhyme duration in different ways. Thus, the peak, expressed in terms of the peak proportion, was affected directly by the right-hand contexts like the word boundary and stress clash, and indirectly by the change of the rhyme duration induced by the word boundary, stress clash and speech rate change.

Caspers and van Heuven (1993) studied the invariant phonetic features of pitch accents ('accent-lending pitch movements') in Dutch, in terms of the shape, the pitch level and the alignment with the segmental string. They used three kinds of time pressure (speech rate, vowel length and tonal composition) 'as an experimental tool for focusing the communicatively important properties of pitch movements' (p. 162), expecting that 'the speaker would have to sacrifice properties of lesser communicative importance while preserving the more essential ingredients as much as possible' (p. 162). They found that, when the vowel was phonologically short, the excursion size for both the rise and the fall increased, the duration of the rise became shorter, and the slope of the rise and fall was steeper. They also found that the shape of the rise was influenced—shortened and steepened—by the presence of a following fall, while the shape of the fall did not change, regardless of the presence or absence of a preceding rise. As for the alignment, the onset of the F0 rise was consistently aligned with the beginning of the syllable, while the offset did not have a fixed alignment point and was shifted earlier mainly because of the closeness of the following accent-lending fall. Overall they concluded that the alignment of the onset is invariant and essential for the rise, while the shape for the fall.

Prieto *et al.* (1995) explored the F0 peak alignment of the H* accent in Mexican Spanish in different structural positions. In the language there are cases in which phrase-medial F0 peaks are displaced further back from the syllable perceptually associated with the H* pitch accent (as shown in Figure 1.1). Using a similar experimental design to Silverman and Pierrehumbert (1990), they attempted to clarify the relationship between the peak location, the segmental duration and other prosodic factors, providing a descriptive model of the F0 peak placement. A test word which has H* on either the initial, medial or

final syllable, was placed under different prosodic environments (intonational-phrase-end, intermediate-phrase-end or phrase-medial), varying the distance between the target stressed syllable and the following stressed syllable. They demonstrated the effects of the right-hand prosodic contexts on the F0 peak alignment such as a word boundary, an intonational- or intermediate-phrase boundary and proximity of the following stressed syllable: the F0 peak was shifted earlier. Furthermore, as the syllable duration increased, the duration from the beginning of the syllable to the F0 peak—what they called 'peak delay'—increased accordingly: the F0 peak occurred later.

All these earlier studies demonstrated that the alignment of F0 targets is influenced by contextual factors—word boundary, phrase boundary, stress clash, tonal crowding and so on.[5] In other words, the F0 peak location is a consequence of the complex interaction between various factors. Particularly, it is obvious that the contextual factors influence the F0 peak alignment in complicated ways. Although the extent of the effects of these factors differs between the languages, this can be basically regarded as avoiding a clashing situation: shifting the pitch accent earlier to put distance, or lengthening the associated syllable to provide more time (or both).

More recently, further evidence on the regularity of tonal alignment has been provided in different languages, particularly by Arvaniti, Ladd and their colleagues (Arvaniti *et al.* 1998; Ladd *et al.* 1999, 2000; Atterer and Ladd 2004; Schepman *et al.* 2006, among others). Overall, it has been revealed that, when the factors affecting pitch accent alignment are properly controlled, both the F0 maxima and minima for pitch accents are consistently aligned with specific segmental landmarks. Moreover, some of the work by Arvaniti, Ladd, and their colleagues share a view that both the beginning and end of a pitch accent are independently anchored at a specific point in the segmental string, which they call 'segmental anchoring', and that these anchored F0 turning points (F0 maxima and minima) mainly contribute to shaping an F0 movement of an utterance. In the following sections, I review such studies about tonal alignment regularities, with a re-examination of the results of the studies discussed above. I also discuss relevant studies on tonal alignment in Mandarin Chinese and in Romance languages.

[5]There is also an extensive investigation into factors affecting tonal alignment in Mandarin Chinese carried out by Xu and his colleagues (Xu 1997, 1998, 1999, 2001; Xu and Wang 2001; Xu and Sun 2002, among others). For example, Xu (1997) provided extensive data about surface F0 variations in different tonal contexts. Since Xu and his colleagues gave their own framework to explain surface F0 variations in Mandarin Chinese, their work is discussed in Section 1.5.

1.3 Regularity of alignment patterns

Arvaniti *et al.* (1998) ran an experiment in Modern Greek to see what factors are responsible for the variability of the F0 peak of prenuclear accent rises, and to evaluate two possible phonological interpretations (a bitonal accent, L*+H, or a phrase H following a pitch accent L*) proposed by Arvaniti and Ladd (1995). They unexpectedly found that, regardless of durational variation by segmental composition, the F0 peak for the prenuclear accent was consistently aligned at a specific point in the postaccentual vowel. This result led them to another experiment in order to confirm this consistent alignment under a condition in which segmental duration was systematically varied. They replicated the results of the first experiment, and found that there was no systematic relationship between the rise duration and the F0 difference between the L and the H targets, which clearly indicated that the L and H tonal targets were independently scaled. They then conducted another experiment to test whether these consistent alignment patterns were unaffected in different prosodic environments, varying word stress location and proximity of the following accent. They found that there was little or no effect of the accent position in a word on the F0 peak alignment, and no effect of the following accent when there were at least two unaccented syllables between the accents. They claim that the F0 targets are independently aligned with a specific point in the segmental string.

One of the important findings from Arvaniti *et al.* (1998) is that both the F0 valley and peak for the prenuclear pitch accent rise is aligned outside the accented syllable. Since, in the autosegmental-metrical theory, a starred tone, either H* or L* regardless of whether it is a monotonal or bitonal pitch accent, is phonetically presumed to occur within the tone-bearing unit, this is an immediate mismatch between autosegmental representation of segments and tones and surface temporal synchronisation between segmental events and tonal events.

Ladd *et al.* (1999) investigated the effect of speech rate (fast, normal and slow) on rising prenuclear pitch accents in Standard Southern British English. Based on the findings of Arvaniti *et al.* (1998), they predicted that 'the tonal targets would be closer together at fast rate and farther apart at slow rate, in correlation with the interval between their segmental anchors'. The results showed a significant positive correlation between syllable duration and F0 rise duration, and also invariant alignment points for both the valley and peak of rising F0 contours in the segmental string regardless of the difference in speech rate. The results thus showed that the interval between the tonal targets varied due to durational change between the landmarks induced by speech rate change. They claimed that the F0

level and alignment of tonal targets are the primary determinants of the shape of a pitch accent.

Ladd *et al.* (2000) explored phonological factors affecting the temporal peak alignment of rising prenuclear pitch accents in Dutch. They conducted an experiment to look into how phonological vowel length affects the alignment of rising pitch accents. They showed different alignment patterns depending on phonological vowel length: the peak of the rises is aligned earlier when the vowel in the accented syllable is phonologically long than when it is phonologically short. They also showed that the F0 valley of the rises was consistently aligned close to the beginning of the accented syllable. They claimed that there are two possible reasons for this result: one is consistent length of the F0 rise; the other is segmental anchoring of the F0 peak to the end of the accented syllable. They performed another experiment to decide between these two accounts, by taking advantage of the fact that phonologically long high vowels in Dutch are phonetically almost as short as that of its counterpart, phonologically short high vowels (Nooteboom and Slis 1972). The result showed that there is a significant difference in alignment between phonologically long vowel and phonologically short vowel, though the alignment of the F0 peak was after the end of the vowel (i.e. the end of the syllable) in both cases. They interpreted the result as follows: the F0 peak alignment is at the end of the vowel for the long vowel and in the following consonant for the short vowel, except for the phonologically long unrounded high front vowel in Dutch which is too short for the rise to complete.

Schepman *et al.* (2006) elaborated on the findings of Ladd *et al.* (2000). Dutch materials were used in which phonological vowel length, and the presence or absence of a suffix and prefix were carefully prepared, in order to explore the effects of phonological vowel length and right-hand contextual factors such as stress clash and word boundary. They found that there is an influence of phonological vowel length on alignment in Dutch which is possibly partly independent of syllable structure. By comparing different dependent variables defining alignment, they also demonstrated that the most appropriate measurements for alignment are time intervals between the F0 target in question and a close segmental landmark.

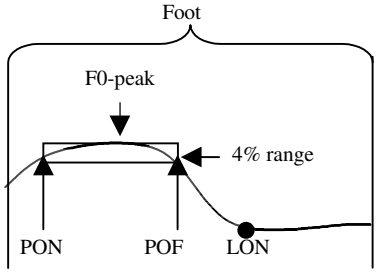
Atterer and Ladd (2004) explored the alignment of the F0 valley and peak for rising prenuclear accents in Northern and Southern German. They collected data to compare the alignment patterns between these German regional varieties, and also these German alignment patterns to those in English and Dutch reported in their other studies (Ladd *et al.* 1999, 2000). The results showed that rising prenuclear accents in German were aligned consistently later than those in English and Dutch, and that, between the German

regional varieties, they were aligned consistently later in Southern German than in Northern German. Atterer and Ladd conducted another experiment to collect data spoken by German learners of English in order to look into whether the alignment patterns found in the first experiment were carried over into the L2 spoken data. They found similar alignment patterns to those found in the first experiment: the rising prenuclear accent of Southern German learners was aligned later than that of Northern German learners, which was, in turn, aligned later than that of native English speakers.

Based on the results of these two experiments, Atterer and Ladd claimed that it is difficult to explain these fine-grained phonetic differences in alignment in terms of phonological association which presumes categorical differences between distinct representations. Rather, they considered these alignment patterns as part of language-specific phonetics based on a phonetic continuum of alignment, comparable to cross-linguistic VOT differences in stop consonants.

There are, to my knowledge, few experimental studies on the regularity of tonal alignment under pitch range variation. Knight (2002) explored the effect of changes in pitch span on the alignment of the F0 peak and plateau.[6] Assuming that it takes more time to reach a higher target in expanded pitch span, she predicted two possible alignment patterns of the F0 peak, and the onset and offset of the F0 plateau. One is that along with later peak alignment both the beginning and end of the plateau may shift later whereby the whole plateau occurs later. The other is that, in order to reach a higher target during the limited amount of time available, the beginning of the plateau may occur later as well as the F0 peak, while the end of the plateau earlier, whereby the whole plateau may be shorter and more tapered. She found that the peak and the onset of plateaux were shifted later due to pitch span expansion. However, she unexpectedly found consistent alignment of the offset of plateaux, regardless of changes in pitch range. Knight suggested that the

[6] A *plateau* is defined as 'the range of times around the peak where the F0 value was within 4% (approximating to a range of perceptual quality)' (House *et al.* 1999). As shown in the figure below, the beginning and end of the plateau are called PON and POF, respectively. LON is the onset of the low level tone.



Adapted from House *et al.* (1999).

stable alignment of the end of plateaux may be of some importance to signal linguistic structure.

1.4 Romance languages

There is also extensive work on tonal alignment in Romance languages recently carried out by various researchers (D'Imperio 2000, 2001; Face 2001; Gili Fivela 2002; Prieto *et al.* forthcoming, among others). Based on the data of three Romance languages, Catalan, Pisa Italian and Neapolitan Italian, Prieto *et al.* (forthcoming) pointed out that these languages have more intricate alignment contrasts than the traditional autosegmental-metrical theory expresses. For example, Catalan has a three-way phonological distinction between rising accents: rising accents with aligned peaks, rising accents with delayed peaks and posttonic rises. In Pisa Italian, there is a difference in the relative position of the peak within the syllable between broad and narrow focuses: while in a broad focus accent the F0 maximum is reached later in the syllable, in a contrastive interpretation the peak is reached earlier within the syllable. In Neapolitan Italian, later alignment of the F0 peak distinguishes a question from a statement.

In order to properly represent the alignment contrasts in these Romance languages, Prieto *et al.* (forthcoming) proposed 'phonological anchoring' which is a development of 'secondary association' originally devised in Pierrehumbert and Beckman (1988). By phonological anchoring, tones of a pitch accent, like a phrase tone, can be secondarily associated with metrical edges such as the end of a mora, a syllable and a prosodic word. They claim that phonological anchoring helps 'clarify the mapping procedure between phonological representation and the surface alignment of pitch accents'.

While phonological anchoring is an engaging refinement to describe alignment differences, it is not free of any complication. As pointed out by Atterer and Ladd (2004), the encoding of alignment differences into phonological representation can lead to an unjustifiable increase of distinctions, as their data clearly display. Atterer and Ladd also allege that it is difficult to treat small alignment differences as phonological categories, and claim that they are best described as a continuum of phonetic alignment. Although both notions of 'phonological anchoring' and 'a continuum of phonetic alignment' intend to account for rather small alignment regularities, there is a disparity between them in terms of how to encode them into a representation.

1.5 Mandarin Chinese

A rather large-scale investigation of tonal alignment in a non-European language is a series of work on Mandarin Chinese by Xu and his colleagues (Xu 1997, 1998, 1999, 2001; Xu and Wang 2001; Xu and Sun 2002). Based on various evidence from their experimental studies, Xu and Wang (2001) proposed a framework for modelling surface F0 contours. In their framework, there are four pitch targets and four rules of implementation. The four pitch targets are two static and two dynamic: [high], [low], [rise] and [fall],[7] and the implementation rules are:

> 1. A pitch target is implemented in synchrony with the host, i.e., starting at its onset and ending at its offset.
> 2. Throughout the duration of the host, the approximation of the pitch target is continuous and asymptotic.
> 3. A falling F0 movement is implemented faster than a rising movement.
> 4. A pitch target containing a high pitch point is implemented with a higher F0 peak when followed by a pitch target containing a low point than when followed by a pitch target with no low point.
>
> Xu and Wang (2001, p. 322)

They also suggest that the implementation rules described above are based on articulartory constraints: 'the maximum speed of pitch change the larynx can produce', 'the maximum speed at which the larynx can change the direction of F0 movement', and 'articulatory coordination of the production of pitch targets and their hosts' (Xu and Wang 2001, p. 329).

One of the characteristics to note about the work by Xu and his colleagues is that it is heavily dependent on data from Mandarin Chinese. This leads them to build their own framework which has fundamental differences from the autosegmental-metrical theory of tone and intonation on which the current study is based. Unlike the autosegmental-metrical theory of tone and intonation, F0 peaks and valleys do not have their own phonological significance in their framework. They are the by-product of pitch targets and their implementation. It is claimed that the F0 peak can be predicted by '(a) the property of the pitch target, (b) the properties of the adjacent pitch targets, and (c) the duration of the

[7] Xu and Wang (2001, p. 321) proposed that the pitch targets are 'the smallest articulatorially [*sic*] operable units associated with linguistically functional pitch units such as tone and pitch accents'. Thus, the tones in their framework are more abstract than those in the framework on which the current study is based.

host' (Xu and Wang 2001, p. 331). Moreover, according to the framework, peak delay is given different explanations. It is supposed to occur under two situations: 'when the pitch target is [rise] and is followed by a target with a low pitch onset'; 'when the pitch target is [high] and surrounded by low pitch values and the duration of the host is sufficiently short' (Xu and Wang 2001, p. 332).

Although various F0 phenomena in Mandarin Chinese are well accounted for in their framework, one of the critical issues to consider is, as mentioned above, that their work relies almost exclusively on Mandarin Chinese data. Since their theory aims to explain surface F0 contours basically in any languages as the auto-segmental metrical theory does, it is important examine which theory is more plausible for further data.

1.6 *Ososagari* and tonal alignment

There is a well-known phenomenon in Japanese called *ososagari* ('late fall') in which the beginning of the F0 fall (i.e. the F0 peak) for a pitch accent occurs after the end of the associated mora (Neustupný 1966; Sugito 1982). While it has been known for some time, there have been only a few studies on ososagari so far.[8] Among the few, Sugito (1982) reported that ososagari tended to occur in the initial-accented words whose second mora had a non-high vowel. More recently, along with the development of the autosegmental-metrical theory of tone and intonation, two small scale experimental studies on tonal alignment were carried out by Kagomiya (1998) and Shinya and Takasawa (1999). Kagomiya (1998) found that the F0 peak for a pitch accent was aligned differently in terms of the accent location of a word. Shinya and Takasawa (1999) demonstrated the effect of right-hand prosodic boundary on the alignment of the F0 peak for a pitch accent, as did some of the early studies on alignment in other languages. Considering the preliminary findings of these studies, ososagari is more likely to occur in initial-accented words (and possibly words with accent on the second syllable), and not in words with accent on other syllables; the F0 peak for a pitch accent may be affected by a right-hand prosodic boundary. On the other hand, it seems fair to say that these studies are far from satisfactory. For example, Sugito (1982) is not a quantitative piece of research since the findings were based on a very small set of samples, and the relevant segmental sequences of the samples included obstruents, which perturbed the F0 tracking, and made it impossible to look into the precise alignment patterns through the target portion. Thus, in the light of consistency in tonal alignment recently found in various languages, it seems obvious

[8]The relationship between F0 peak location and accent *perception* has been relatively more thoroughly studied (Sugito 1982; Hasegawa and Hata 1992).

that quite a few questions remain unclear about tonal alignment, including ososagari, in Tokyo Japanese.

One of the main focuses of the current study is on the alignment of the accentual F0 rise at the beginning of the initial-accented word. Based on Venditti (2005), which is a revised version of Pierrehumbert and Beckman (1988), the tonal structures of interest in the current study are as follows:[9][10]

(1.2)

utterance-initial	%wL	H*+L ...
utterance-medial	wL%	H*+L ...

There are two main reasons for using this specific LHL sequence in Tokyo Japanese. Firstly and most importantly, this is *the* tonal sequence in Tokyo Japanese which is most comparable to those used in previous studies on tonal alignment in other languages.[11] The data on this tonal sequence thus enables us to make a direct comparison with the alignment patterns between the languages. Secondly, as reported in previous work (e.g. Sugito 1982), ososagari is more likely to occur in initial accented words, and the segmental composition of the accented syllable also seems to play a role in the F0 peak location. Considering the findings from previous work, I presume that the data of initial-accented words are most crucial to revealing tonal alignment in Tokyo Japanese. In addition, it is, practically, much easier to prepare materials with exhaustive segmental combinations in an initial-accented syllable than in a non-initial accented syllable.

Before presenting the main purposes of the experiments of the present study, I give a summary of Tokyo Japanese prosody in the following two sections. The next section provides a brief outline of accent and intonation in Tokyo Japanese. The section after the next discusses, with various types of evidence, the phonological relevance of syllable, mora and quantity in Tokyo Japanese, relevant to the current study.

[9] wL is a variant of the phrase-initial low tone. It is posited to describe the shorter and higher-valued F0 valley occurring at the beginning of the accentual phrase with accent on the initial syllable and with the long initial syllable (Pierrehumbert and Beckman 1988).

[10] Although the associations of the boundary low tones (%wL and wL%) are different ('%' indicates the location of the accentual phrase boundary with which the boundary tones are associated), these two structures are equivalent: the boundary low is at the (left or right) edge of the accentual phrase, and the pitch accent, H*+L, is associated with the accented initial syllable.

[11] Of course, the pitch accent type of the accentual rise of the current study is not identical to that of the previous studies. For example, H* was the pitch accent in Silverman and Pierrehumbert (1990) and Prieto *et al.* (1995). In some other studies, the *pitch accent type* is simply their empirical and theoretical question. Nonetheless, they are all an accentual F0 rise.

1.7 Accent and intonation in Tokyo Japanese

While the accent and intonation system of Tokyo Japanese has been studied in a variety of frameworks, there are several points generally agreed across them:

(1.3)
a. One characteristic pitch pattern, namely a high-low tonal sequence, marks the word accent.
b. A word has at most one accent on any syllable or can be unaccented.
c. Thus, *n*-syllable words have $n + 1$ possible accentuations.
d. Phrase initial moras have a low tone and second moras have a high tone unless the word in that position has an initial accent.

Keeping these in mind, I will outline the ways in which different analyses describe the pitch accent system of Tokyo Japanese in the following paragraphs.

In the traditional description of word accent in Japanese, it has been assumed that each mora in a word is lexically specified as either High or Low, and this description is still common in quite a few dialectal studies (e.g. NHK 1998). A classic generative description by McCawley (1968) also treats word accent in the same fashion.

(1.4)

ka ra su	ko ko ro	mu su me	sa ku ra
'crow'	'mind'	'daughter'	'cherry tree '

Words in (1.4) are a set of examples of three-syllable words. As stated in (1.3c), there are four accent types. The last two appear to have the same accent pattern as a citation form, but the former has the accent on the third syllable, while the latter has no accent. With the nominal particle, *-ga*, attached after the words, the accentual difference becomes clear, as shown below.

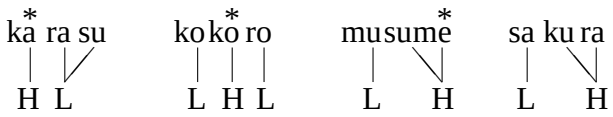
(1.5)

mu su me (-ga) sa ku ra (-ga)

In an early autosegmental analysis of Japanese accent by Haraguchi (1977), tones and vowels are separated on different tiers ('segmental' and 'tonological'), and, when a word is an accented word, only one vowel of the word is specified with a star (*) in the lexicon. In a series of derivations, tones (H and L) are inserted and fully associated with the vowels

by various association rules and conventions. Applications of them generate surface accent patterns, comparable to those in (1.4), as in (1.6).

(1.6)



While Haraguchi's work seems to employ richer representation and more elaborated tonal specification mechanisms, it is little more than a notational variant of the traditional analysis. Considering the results from more recent studies, there is no distinction between the H tone associating with the accented mora and the H tones associating with the unaccented moras, in spite of the fact that they are phonetically different. Moreover, like the traditional description, his analysis assumes fully specified tonal representation.

An essentially different approach to Japanese accent and intonation has been developed over the last two decades, based on a series of experimental studies (Poser 1984; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Venditti 1997, 2005). This approach, together with intonation studies of other languages, is now called the autosegmental-metrical theory of intonation (Bruce 1977; Pierrehumbert 1980; Ladd 1996, among others). Unlike previous theories which assume full association of tones with tone-bearing units, this theory regards tones to be sparsely distributed at well-defined locations in the prosodic structure (*phonetic underspecification*). A distinction is made between the pitch accent, H*L, and the phrasal high tone, H, which are not treated as different elements in previous studies such as McCawley (1968) and Haraguchi (1977), as discussed above. An example of the structure is shown in Figure 1.2. A pitch accent, HL, associates with a specified mora (e.g. /se/ in /seetaa/). A boundary low tone, L, and a phrasal high, H, associates with the first mora /a/ and the second mora /ne/ of an unaccented word /ane/, respectively. The boundary tone, L, marking the right edge of the accentual phrase /ane-no/ associates with the first mora of the following accentual phrase /akai/ since the first syllable is short and unaccented.

1.8 Syllable, mora and quantity in Japanese

Different languages exploit possible prosodic units to different degrees in their own ways. In Tokyo Japanese, the syllable and the mora are both relevant prosodic units below the prosodic word for the description of the language.[12] On the one hand, much evidence is

[12]The (bimoraic) foot is also an important unit below the prosodic word, but it is not considered here.

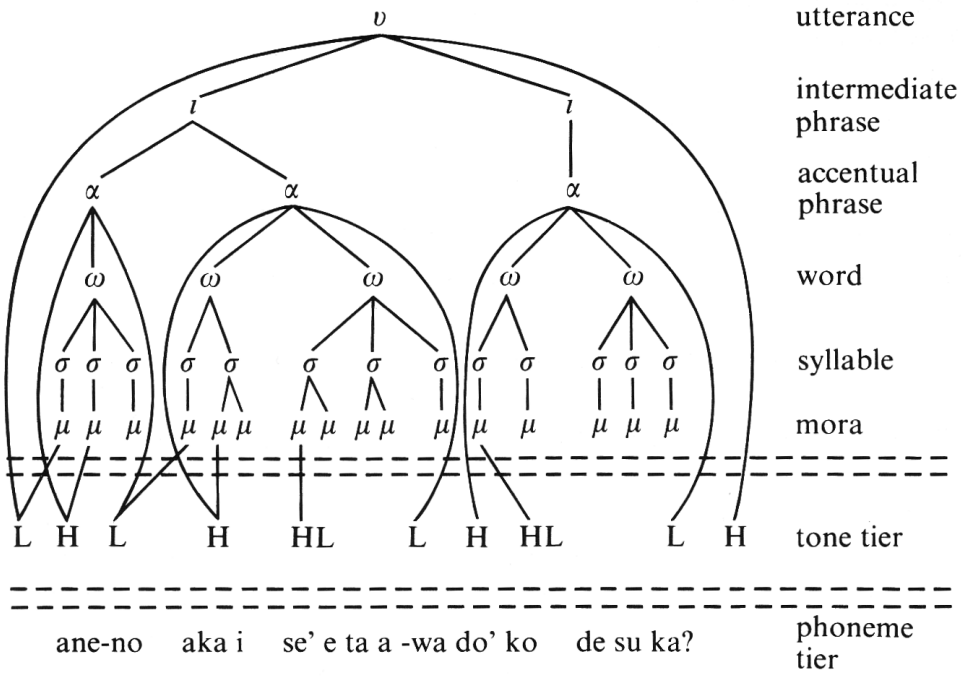


Figure 1.2: Prosodic and tonal structure of a phrase /ane-no akai se'etaa-wa do'ko desuka/ 'Where is big sister's red sweater?' adapted from Pierrehumbert and Beckman (1988, p. 21)

provided that the mora plays a role in various phonological and morphological phenomena such as a timing unit, a determinant of phonological length of words and phrases, the segmentation unit, the perceptual unit, and the unit defining the bimoraic foot. On the other hand, although it has been neglected in Japanese linguistics for some time, the syllable is also an indispensable unit to characterise phenomena related to word formation and syllable weight.[13]

While various kinds of evidence support the significance of the mora and the syllable in Tokyo Japanese, the definitions of the tone-bearing unit and the accent-bearing unit, where the mora and the syllable are arguably relevant, are implicit and inconsistent across the analyses. In a linear phonological analysis by McCawley (1968) where fully tonal specification for the mora is presumed, it is argued that both the mora and the syllable are relevant for accentuation: the former as the tone bearing unit, and the latter as the accent bearing unit. Given that each mora is fully associated with a tone, one-syllable word,

[13]For a comprehensive review of phonological relevance of the syllable and mora in Japanese, see Kubozono (1999).

/kaN/ (HL) ‘completion’ has a High on the first mora, /ka/, and a Low on the second mora, /N/, while /kaN/ (LH) ‘sense’ has a Low on the first mora and a High on the second mora. On the other hand, the syllable also plays a role in accentuation. It is regarded as the bearer of accent, since a two-mora syllable always has a accentual High on the first mora. For example, while /koorogi/ ‘cricket’ (HLLL) is an actual word in Tokyo Japanese, there is no such word as /koorogi/ (LHLL). However, /kogaisja/ ‘subsidiary company’, whose second mora is a syllable itself, can bear an accentual High on the second mora with the accent pattern as LHLL (Shibatani 1990, p. 160). If the mora alone determined Japanese accentuation, then this restriction cannot be invoked. A clearer case is the loanword accentuation. Loanwords can be divided on the basis of their accentuation. One of the groups follow the ‘loanword accent rule: place an accent on the syllable containing the antepenultimate mora’ (Kubozono 1999).

(1.7) a. o-o.su.to.ra.ri.a ‘Australia’, de-n.ma-a.ku ‘Denmark’
b. wa.si-n.to-n ‘Washington’, su.pa-i.da-a ‘spider’

(In (1.7), the dot and the hyphen are syllable and mora boundaries, respectively. Where there is a syllable boundary, there is also a mora boundary. The location of lexical accent is underlined. Examples are taken from Kubozono (1999).) While, with words like those in (1.7a), it appears that the accent is placed on the antepenultimate mora, it is necessary to refer to the syllable containing the antepenultimate mora to explain the accent placement of words like those in (1.7b).

Although it is certain that the syllable plays a role in accent placement, it is imprecise to define it as ‘the accent-bearing unit’ in the theory where full association is assumed. As stated in (1.3a), the high-low tonal sequence marks the word accent, and, in an accented word with only short syllables like *ko.ko.ro* ‘heart’ (LHL), the last two syllables (‘*ko.ro*’) bear the accent (HL), which is inconsistent with the definition that the syllable is the accent-bearing unit.

In Haraguchi (1977), the tone-bearing unit is assumed to be vowels and the moraic nasal, both of which, he argues, should be voiced. It is implicitly regarded that the star (*) is specified to the first mora of two-mora syllables like vowel geminates or vowels followed by a moraic nasal. Since the specification of the accentual HL sequence is determined by association rules and conventions, the syllable is irrelevant in marking the accent.

In the autosegmental-metrical theory of Japanese intonation, Poser (1984) proposes that the accent is a H tone linked to a particular mora in the lexical entry for accented words.

Pierrehumbert and Beckman (1988) propose that the accent tone, T, on the tone tier which is linked to the mora or the first mora of a two-mora syllable branches into the H and L, while they do not clearly state whether accentual tones are lexical or post-lexical.

1.9 Outline of the thesis

Three speech production experiments were performed in the current study. Chapter 2 gives an overview of the data collection. Chapters 3, 4 and 5 present the data collected under different experimental conditions. The first experiment examined the alignment of the F0 targets at the beginning of initial-accented words, varying the syllable/mora structures of the accented syllable. The second experiment explored how the alignment patterns, found in the first experiment, were influenced in different speaking modes; the speaking modes of interest were fast speech rate, raised voice, and local emphasis. The third experiment compared the F0 peak alignment of unaccented and non-initial-accented words to those of initial-accented words. In Chapter 6, in the light of the data gathered in these experiments, I discuss theoretical issues outlined in Chapter 1, and also suggest future directions of research.

CHAPTER 2

General Method

2.1 Introduction

This chapter provides an overall description of the data collection and analyses of this project. As stated at the end of the previous chapter, there are three chapters on the experiments, and each chapter is concerned with different factors of interest related to tonal alignment:

Chapter 3 syllable and mora structures
Chapter 4 speaking modes
Chapter 5 accent types and locations

The data were collected via a series of recordings: it took approximately an hour for a speaker to perform. Each session has an analogous arrangement, apart from the factors in question. The details of the recordings are first described in the following section.

2.2 Materials

2.2.1 Test words

Fifteen groups of five test words were prepared in such a way that each group varied in initial segmental composition and accent type (see Tables 2.1, 2.2 and 2.3). Test words were systematically collected to contain only vowels and sonorants (mostly nasals) in target sequences for the F0 tracking and segmentation, and to balance, as far as possible, the effects of intrinsic segmental properties on the F0 in each group. While the number of moras of unaccented and initial-accented words ranged between 3 and 5 (mainly

3 or 4), that of second-, third- and fourth-syllable accented words was between 3 and 6 (mostly 4 or more). The target sequence of the test words was the first two moras for unaccented words, and the accented mora and the mora following it for accented words. Six different types of initial segmental make-up were arranged only for unaccented and initial-accented words, as shown in Tables 2.1 and 2.2. Second-, third- and fourth-syllable-accented words were less comprehensive and thus fewer than unaccented and initial-accented words, because of the limitations of collecting possible real words (see Table 2.3).

#CV+CV…	ne+mimi	ni+meN+sei	ni+mame	ni+mono	ma+na+mi
#CVCV…	mimi+nari	minamoto	mono+no+ke	nami+nari	nama+nie
#CVR…	mee+moku	moo+moku	noo+miN	neemiNgu	maazjaN
#CVV…	mainasu	nai+meN	moe+nokori	mae+muki	niamisu
#CVN…	maNneri	niNniku	niN+mei	neN+matu	naN+miN
#CVQ…	maQ+seki	maQ+satu	miQ+situ	niQ+saN	meQki

Table 2.1: Unaccented words (# is a word boundary, and + is a morpheme boundary. C is an onset consonant, and V is a vowel. R, N and Q are the second part of a long vowel, a moraic nasal and moraic obstruent, respectively. The second vowel of CVV is different from the first one. (i.e. 'VR' indicates a long vowel, and 'VV' of CVV a diphthong.)

#CV+CV…	**mo**+naka	**na**+no+hana	**no**+naka	**no**+miti	**mi**+nari
#CVCV…	**mi**nimamu	**mo**mizi	**na**mida	**mi**neraru	**me**morii
#CVR…	**naa**basu	**nuu**doru	**mee**+nichi	**noo**+nai	**mii**ra
#CVV…	**mai**+nichi	**mai**rudo	**noi**ze	**nao**+ki	**noe**ru
#CVN…	**meN**ma	**maN**mosu	**neN**+maku	**niN**+mu	**neN**ne
#CVQ…	**meQ**seezi	**meQ**ka	**maQ**+ki	**maQ**to	**miQ**to

Table 2.2: Initial-accented words (see Table 2.1 for special symbols).

2.2.2 Carrier sentences

All the test words were placed in two types of the carrier sentences in the first three sessions:

- *X-ga kaitearimasu.* ('X is written.');
- *Sokoni X-ga kaitearimasu.* ('X is written there.').

There were two reasons for using these two types of carrier sentences; one was to look into the effect of prosodic boundaries (left-hand in this study); the other to manifest the

#CV**CV**CVCVCV…	no**mi**+ja	no**mi**+mizu	ma+**mi**nami	mo**mo**+niku	na**ma**+mono
#CVCV**CV**CVCV…	mame+**mo**naka	mono+**mo**rai	nomi**nee**to	mini+**mo**naka	mina**mi**+mati
#CVCVCV**CV**CV…	minami+**mo**mizi	nama+ma**me**+moti	minami+**mo**naka	minami+**ma**turi	nami+mo**mo**+niku

Table 2.3: Second-, third- and fourth-syllable accented words. The syllable in bold is lexically accented.

F0 valley before the target F0 peak. Additionally, an unaccented accentual-phrase *sokoni* was used in order to avoid a clashing effect due to the use of an accented accentual phrase.

In the terminology of the Japanese ToBI system (Venditti 2005), these two sentences were expected to be pronounced as a single intonation phrase.[1] The first sentence consists of one or two accentual phrases with a test word at the beginning, and the second of two or three accentual phases where a test word was at the beginning of the second. Their tonal structures can be assumed as follows: for the intonation phrases with the target unaccented accentual phrase,

- `%L H- L%`
  One accentual phrase; utterance-initial test word
- `%L H- L% H- L%`
  Two accentual phrases; utterance-internal test word

and, for the intonation phrases with the target accented accentual phrase,

- `%L (H-) H*+L L%`
  One accentual phrase; utterance-initial test word
- `%L H- L% (H-) H*+L L%`
  Two accentual phrase; utterance-internal test word

In the scripts of the fourth session, test words were put in slightly different carrier sentences to help the subjects to emphasise the target word;

- *Y-deha naku, X-ga kaitearimasu.* ('Not Y, but X is written.');
- *Y-deha naku, sokoni X-ga kaitearimasu.* ('Not Y, but X is written there.').

Although preceded by an extra phrase, the target phrases were the same as those of the other three sessions (however, their tonal structures were expected to be different due to the difference in the reading tasks. See Section 3.2.3 for details).

2.3 Subjects

Eighteen native speakers of Tokyo Japanese were recruited at Sophia University in Tokyo. They were students, undergraduates and postgraduates, at the university, aged between 19 and 28: eight female undergraduates, two male undergraduates, one female postgraduate

[1] Comparable to an 'intermediate phrase' in Pierrehumbert and Beckman (1988).

and seven male postgraduates. For their participation in the experiment, the speakers were rewarded with 2000 yen, roughly 10 pounds.

2.4 Recording and reading tasks

The recording was conducted in a sound treated studio at the Phonetics Laboratory of Sophia University in Tokyo, with a digital audio tape (DAT) recorder (Sony TCD-D8) and an electret condenser microphone (Sony ECM-MS907). The recording of each speaker was made up of four different sessions (see below), with a break between each. In each session, after an appropriate instruction by the author and a brief practice, speakers read the sentences printed on sheets of A4 paper, one by one. Test sentences were mixed with filler sentences in random order by the 'rline' programme[2] set in the UNIX/Linux system at the Edinburgh University Department of Theoretical and Applied Linguistics. The first and last sentences of each sheet were filler sentences to reduce the effect of turning sheets. Each sheet had seventeen sentences. The sessions were set up in such a way that the speakers would be unable to tell when a session would finish: filler sheets were placed after the genuine sheets. No instruction was given during the sessions, and misread and disfluent sentences were repeated according to each subject's own decision.

Each session was spent on one type of reading task, and subjects were minimally instructed by the author in what manner to read the material (apart from the instructions below, the subjects were not further directed either before or during the sessions).

First session to read as naturally as possible.
Second to read as if talking to someone on the other side of a busy street.
Third to read as fast as possible.
Fourth to place emphasis on the word printed in bold face.

With these instructions, they read aloud the material, from which four types of data were expected (with a certain degree of intra- and inter-speaker variation): normal, raised voice, fast and local emphasis.

[2] 'rline' is a command-line program which arranges the lines of an input text file in random order, and produce an output text file with the arranged lines.

2.5 Analysis procedures

2.5.1 Corpus description

In anticipation of unusable or unsuitable data, more subjects were collected than necessary for the statistical analyses carried out in this study. Out of the eighteen speakers recorded, complete acoustic analyses were carried out for seven. The rest were discarded for various reasons: (1) most importantly because they found it difficult to speak fast or to raise their voice; (2) because they had lived for a certain period during their lives outside the region where Tokyo Japanese is spoken; (3) because they produced utterances with prosodic structures inappropriate to this study. These seven speakers whose recordings were fully analysed consisted of four females and three males, and none of them had any speaking or hearing difficulties.

2.5.2 Data conversion, annotation and measurements

With the equipment in the TAAL Laboratory of the University of Edinburgh, recorded materials taken on DAT tapes were digitised at a sampling rate of 16 kHz after appropriate low-pass filtering, and transferred to a Sun Sparc Ultra-1 workstation. Digitised speech signals, which were very large files containing many utterances, were cut into small sound files including one test word each, with the aid of Praat scripts written by Mietta Lennes which automatically set boundaries at pauses on the basis of an intensity analysis[3]

Annotation and acoustic measurements were performed using Praat.[4] The segmentation was carried out manually on the basis of the visual display of the oscillogram and the spectrogram with a controlled cursor. If there were two candidates for a segmented point, the earlier was always taken. The segmentation points were basically marked at zero in amplitude (zero crossing), except for cases like the release of voiced stops. Since almost all target sequences were composed of nasals and vowels, there was little difficulty in locating a boundary between the target segments. Only when it was hard to place a boundary in the waveform, did I rely on the wide-band spectrogram. Figure 2.1 shows examples of the segmentation. Target F0 minima and maxima were located via parabolic

[3]For more detail, see `www.helsinki.fi/~lennes/praat-scripts/`
[4]For detail of Praat, visit `www.praat.org`

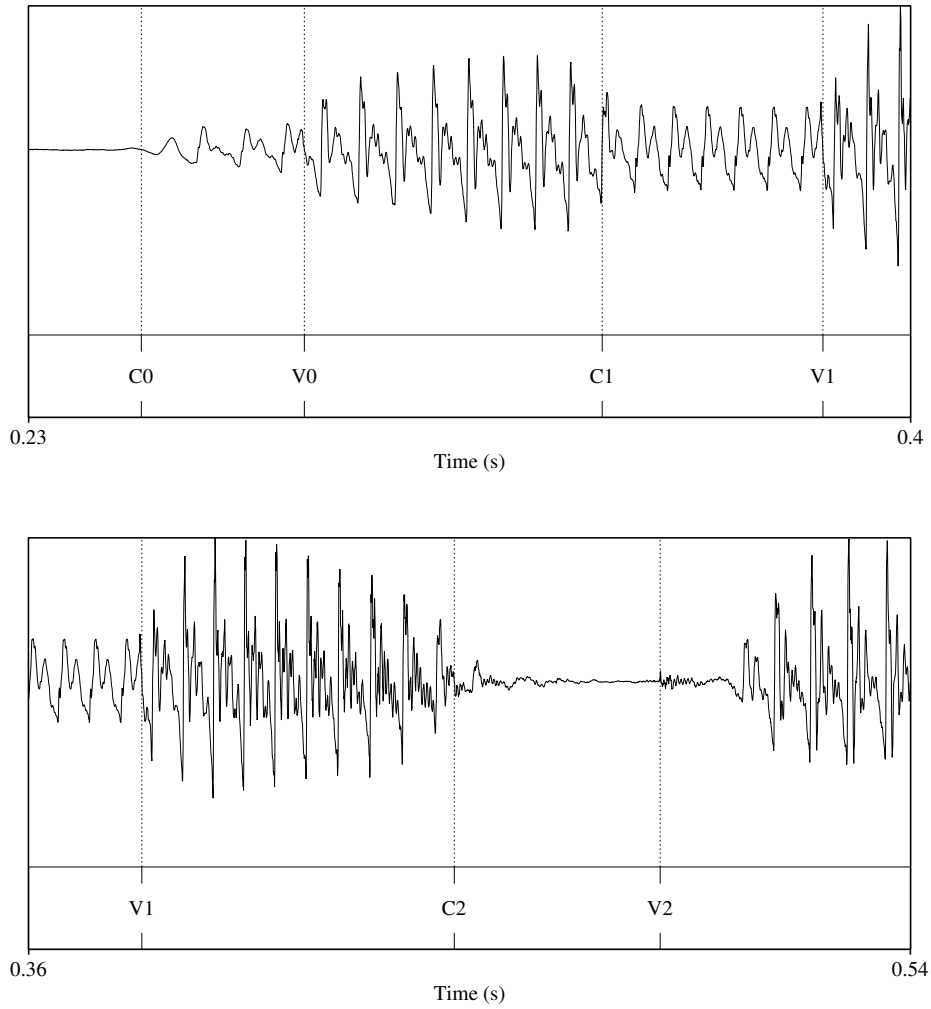


Figure 2.1: Portions of the oscillogram of utterance-initial ‘nonaka’: the top shows /non/ of /nonaka/ (from C0 to V1); the bottom, /ak/ (from V1 to V2).

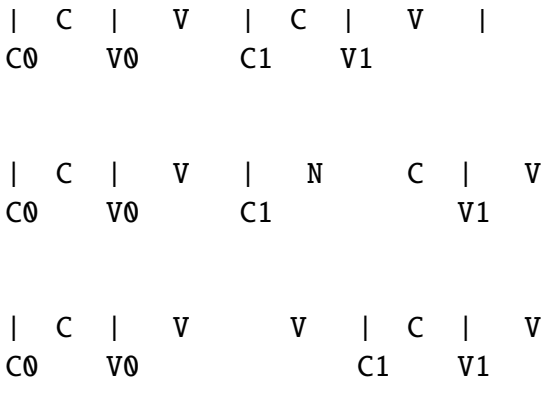


Figure 2.2: Labelling scheme of target sequences. A vertical line ('|') is a segmented point. The letters below the vertical line are the labels. The first segment of a target sequence starts from zero (e.g. 'C0'). The first sequence is for CV+CV and CVCV; the second for CVN; the third for CVR (a long vowel) and CVV (a diphthong).

interpolations between selected portions around them.[5]

Figure 2.2 shows the labelling scheme of target sequences. Target sequences were labelled at the beginning—i.e. point, not section—of each segment, and the first segment of a target sequence was labelled starting from zero. So the *beginnings* of the first consonant and the first vowel of a target sequence were labelled 'C0' and 'V0', respectively. While C0 and V0 can be used as a phonologically common segmental point for the F0 peak alignment (C0 to H, and V0 to H), it was unavoidably impractical to establish a segmental point which was phonologically common to all types of the segmental sequences, if required to choose the nearest segmental point to the F0 peak, as suggested in Atterer and Ladd (2004). For example, 'C1' is acoustically at the end of the initial vowel across the groups, but it is phonologically not the identical place. 'C1' is the end of both the syllable and mora for CV+CV and CVCV; only the end of the mora for CVN; only the end of the syllable for CVR and CVV. Therefore, three types of measurements were used in the following section to examine the F0 peak alignment: alignment of H (F0 peak) relative to C0; alignment of H relative to C1; alignment of H relative to V1.

[5]It might be possible to claim that, instead of the F0 peak and valley, some other features of the F0 curve can be employed as appropriate mesurements. However, there are at least two good reasons for measuring the F0 peak and valley for the current study. Firstly, it is empirically well-established in Japanese phonetics to measure F0 peaks and valleys as linguistically relevant intonational features (e.g. Pierrehumbert and Beckman 1988). Secondly, it often happens that it is unattainable to detect relevant F0 peaks and valleys, thus alternative F0 features need to be drawn on. However, it is of little difficulty to locate the relevant F0 peak and valley in the current study.

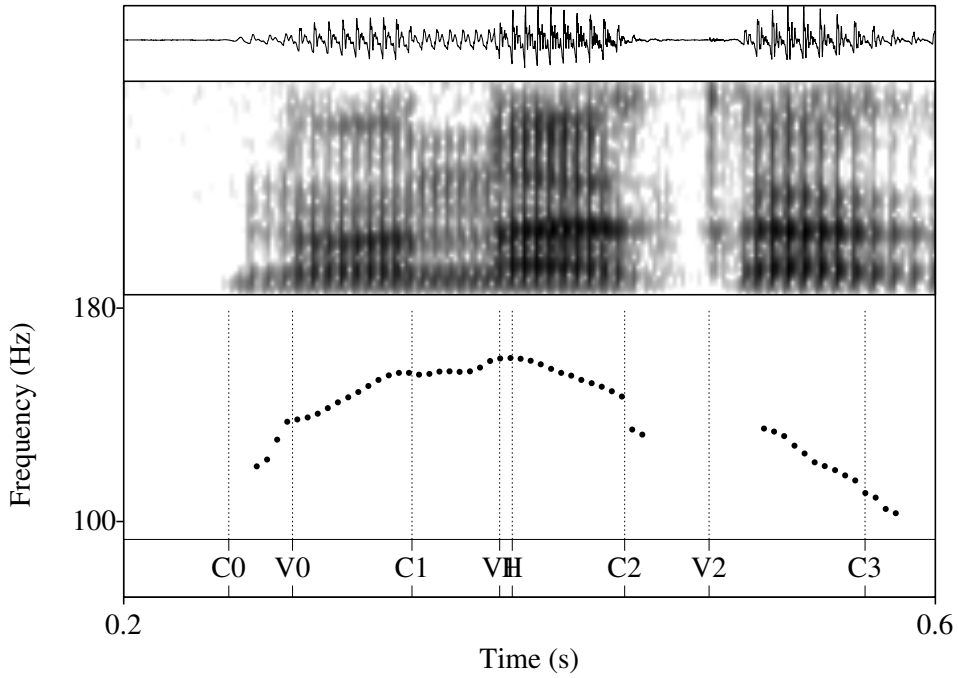


Figure 2.3: Oscillogram, spectrogram, F0 trace and labels of one of the tokens (utterance-initial 'nonaka').

Figure 2.3 is an example of the labelling. As shown in Figure 2.3, both segmental and tonal labels were put in the same field for later measurements. In order to reduce errors of manual labelling, a Praat script was used to semi-automatically place labels at appropriate boundaries in the field, based on the character string of the file name. For example, after manually placing boundaries in the field of a file (e.g. /nonaka/), based on the character string of the file name ('n', 'o', 'n', 'a', 'k' and 'a'), the script was run to insert 'C0' at the beginning of /n/, 'V0' at the beginning of the /o/, and so on. Furthermore, annotated files were later processed with a *bash* shell script which checked missing labels in the annotated files to find mistyping and mislabelling.

After the labelling, semi-automatic data extraction was performed on the duration of portions of interest and the annotated F0 maxima and minima, with the aid of Praat scripts. The values of the annotated F0 minima and maxima were obtained via an autocorrelation method of F0 estimation (Boersma 1993).

2.5.3 Statistical analysis

The statistical analyses consisted of testing for significant effects of independent variables, such as structure and speaking styles, on various dependent variables, mainly using appropriate ANOVAs. The details will be given where relevant in the subsequent chapters.

CHAPTER 3

Regularity in tonal alignment

3.1 Introduction

As described in Chapter 1, there is a well-known phenomenon in Japanese called *ososagari* ('late fall') in which the beginning of the F0 fall (i.e. the F0 peak) for a pitch accent occurs after the end of the associated mora (Neustupný 1966). There are a few preliminary descriptions of ososagari provided so far. For example, Sugito (1982) reported that ososagari tends to occur in the initial-accented word whose second mora has a non-high vowel. However, there has been no detailed study which thoroughly looked into, for example, in what environment ososagari occurs in terms of different prosodic structures, or to what extent it occurs whenever it occurs. It thus seems fair to say that the description of ososagari is still partial and far from complete. One of the purposes of this chapter is to provide more comprehensive data of ososagari (and non-ososagari).

Another purpose is to investigate whether there is any regularity in the alignment of F0 targets, particularly of the F0 peak, in Tokyo Japanese. Recent alignment studies, as discussed in Chapter 1, provide evidence for fairly consistent language-specific alignment patterns of F0 targets (both F0 valley and peak) in various European languages (Arvaniti *et al.* 1998; Ladd *et al.* 1999, 2000; Atterer and Ladd 2004, among others). I would like to examine whether similar alignment regularity can be found in Tokyo Japanese.

The other purpose, related to the second one, is to decide what measurement is most suitable for the current study on tonal alignment in Tokyo Japanese. Different measurements are used in previous studies to explore tonal alignment—for example, the alignment of the F0 peak relative to the beginning of the syllable rhyme, or the alignment of the F0 peak relative to the end of the syllable—mainly because of the different purposes of those

	Phoneme	Syllable	Mora	Gloss
CV+CV	/mi+nari/	**mi**.na.ri	mi-na-ri	‘appearance’
CVCV	/namida/	**na**.mi.da	na-mi-da	‘tears’
CVN	/maNmosu/	**maN**.mo.su	ma-N-mo-su	‘mammoth’
CVR	/nuudoru/	**nuu**.do.ru	nu-u-do-ru	‘noodle’
CVV	/mairudo/	**mai**.ru.do	ma-i-ru-do	‘mild’

Table 3.1: Examples of the five types of the syllable/mora structure for the initial-accented word. ‘C’ is an onset consonant, and ‘V’ is a vowel. ‘N’ and ‘R’ are a moraic nasal and the second part of a long vowel, respectively. The second vowel of CVV is different from the first one; ‘VV’ of CVV indicates a diphthong and ‘VR’ a long vowel. The accented syllable is shown in bold type. A plus sign, a dot, and a hyphen are a morpheme, syllable and mora boundaries, respectively.

studies. It is not yet clear which measurement is more plausible among the different measurements, or whether it should be decided depending on the nature of data to be studied.[1]

As demonstrated in previous alignment studies on other languages, tonal alignment can naturally be a rather complicated subject to deal with because of the large number of factors involved. Some of the factors therefore should be explored separately in order to gain a precise understanding. Considering the findings discussed in Chapter 1, particularly the observation in Sugito (1982) which suggested the relevance of structural difference of the accented syllable, I first set up an experiment to collect data in terms of the different structures of the accented syllable of the initial accented word. There are five different syllable/mora structures of the target accented syllable for the material in this experiment.[2] Sample words for the structures are shown in Table 3.1. The initial syllable bears the accent, and is mono-moraic for the words of CV+CV and CVCV, and bi-moraic for those of CVN, CVR and CVV.

In order to obtain data comparable to those in previous studies, test words were placed in two positions in an utterance, expecting to be produced with tonal structures as shown below:[3]

(3.1) `utterance-initial %wL H*+L ...`
`utterance-medial wL% H*+L ...`

[1] But see the discussions of Appendix A in Atterer and Ladd (2004) and of Schepman *et al.* (2006).
[2] See Section 2.2.1 in Chapter 2 for the full details of the material.
[3] See Section 2.2.2 in Chapter 2 for full details of the carrier sentences.

The H* of H*+L is associated with the initial syllable of the test word. The %wL and wL% are associated with the left and right edges of the acccentual phrase, respectively, without the secondary association with a mora at the accentual phrase boundary.

With the material presented here, the first experiment was carried out in order

- to provide a comprehensive data of the alignment of the F0 targets (including os-osagari) at the beginning of the initial-accented accentual phrase; and
- to test the hypothesis that there are differences in alignment due to syllable/mora structures.

3.2 Results

I start with a visual inspection item by item in Section 3.2.1 in order to obtain a picture of how ososagari occurs in different syllable/mora structures, and also to look into which measurement is suitable for quantitative analyses in the following section on the F0 peak alignment (Section 3.2.2). I then present the data of the alignment of the F0 valley for L: the visual inspection in Section 3.2.3, and the quantitative analyses in Section 3.2.4. Graphical acoustic representations for the visual inspections below are based on the annotated data described in Section 2.5.2 of Chapter 2.

3.2.1 Visual inspection of ososagari and the F0 maximum

Figure 3.1 shows examples of the F0 peak for initial-accented words which begin with the CV+CV and CVCV sequences.[4] Ososagari (i.e. peak delay) clearly occurred in most items across the subjects regardless of the presence or absence of a morpheme boundary.[5] As can be seen in the examples on the left (/na+no+hana/, /mi+nari/ and /mo+naka/), the F0 peak occurred after the end of the accented syllable (i.e. after the end of the vowel of /na/, /mi/, and /mo/, respectively), and was aligned just around the beginning of the vowel of the following syllable, which corresponds to the observation in Sugito (1982) that ososagari tends to occur in initial-accented words whose second mora has a non-high vowel. However, similar alignment also occurred in the initial-accented words whose second mora has a high vowel, as can be seen in the examples on the right. In fact, this alignment pattern was uniformly observed regardless of the segmental composition of the first two syllables, which clearly departs from Sugito's characterisation. I believe that her

[4]For comparison, all the examples in this section are from the data of one male speaker (Speaker TN).

[5]There were eight non-ososagari items out of 140 (20 items $\times$ 7 subjects): four from one of the subjects, and four from another.

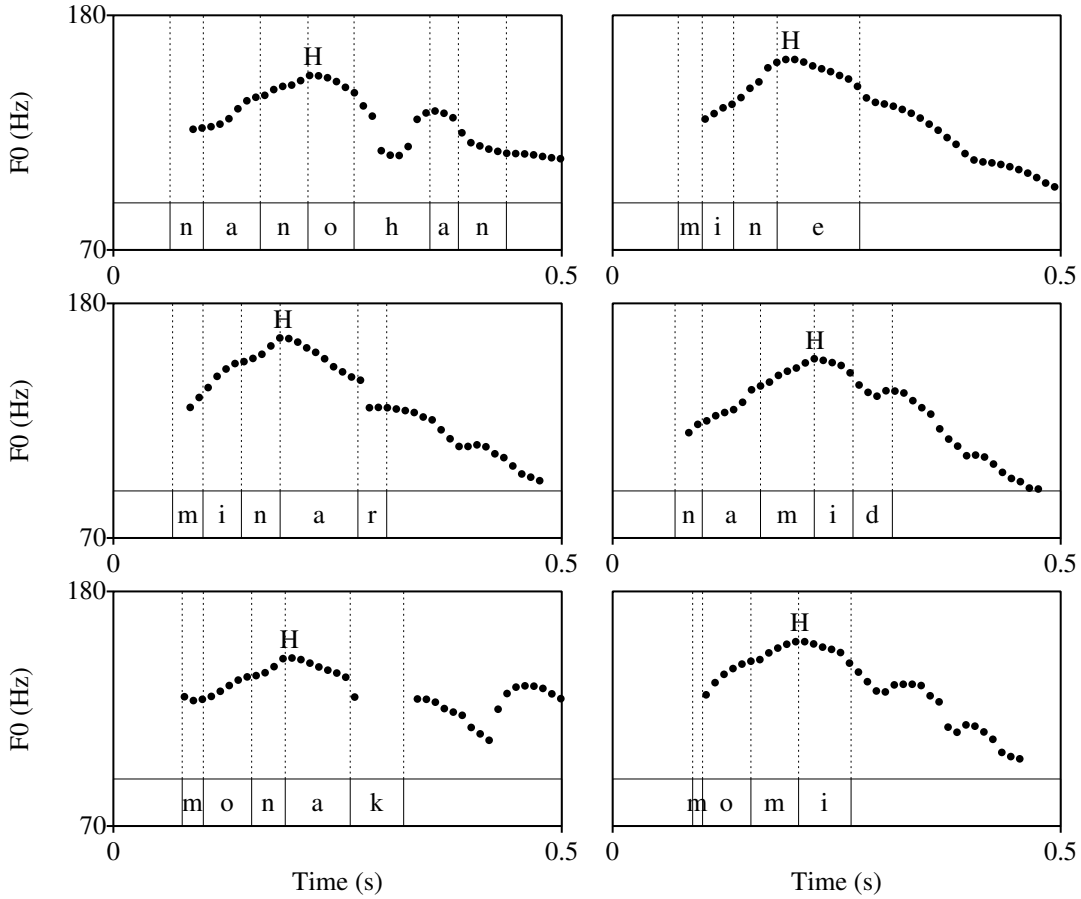


Figure 3.1: Examples of the F0 peak for initial-accented words which begin with the CV+CV and CVCV sequences. The CV+CV items are from the top to bottom on the left: /na+no+hana/ ‘rape blossoms’, /mi+nari/ ‘appearance’ and /mo+naka/ ‘bean-jam-filled wafers; the CVCV items from the top to bottom on the right, /mineraru/ ‘mineral’, /namida/ ‘tears’ and /momizi/ ‘red leaves’.

characterisation is imprecise because it was based on a very small set of words without quantitative analysis.

Clear ososagari did not happen in the items of the other three sequences. In the items of CVN, as shown in Figure 3.2, the F0 peak occurred just around the end of the first mora (around C1 in Figure 2.2 of Chapter 2) in most cases across the subjects regardless of the vowel type of the accented syllable.

In the items of CVR (‘R’ is the second part of a long vowel), the F0 peak occurred within the long vowel of the accented syllable, as shown in Figure 3.3. Although it

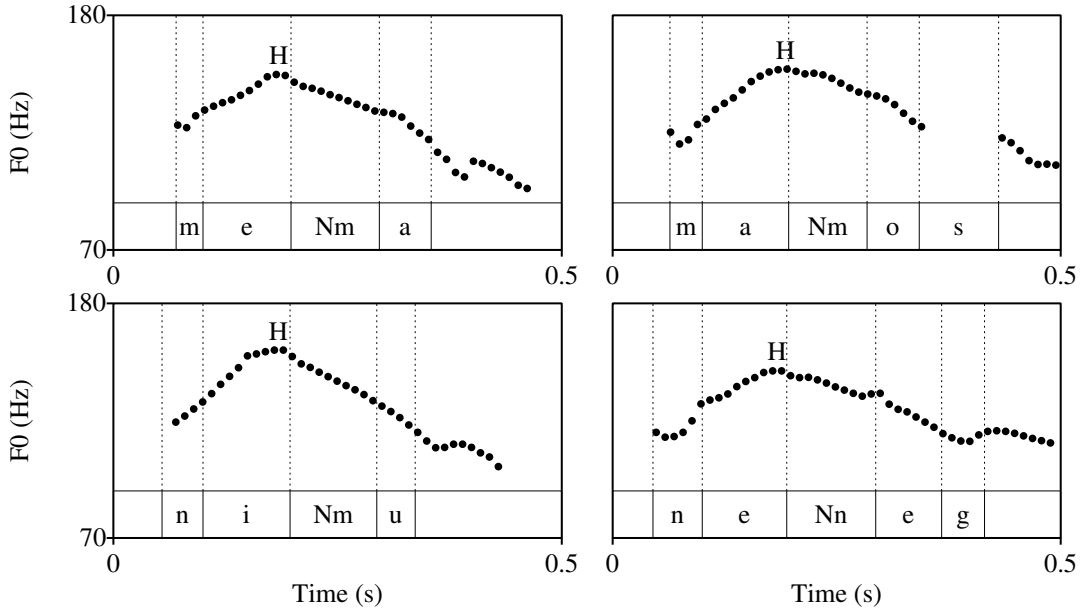


Figure 3.2: Examples of the F0 peak location for initial-accented words which begin with the CVN sequence. /meNma/ ‘bamboo shoots’, /niN+mu/ ‘duty’, /maNmosu/ ‘mammoth’ and /neNne/ ‘sleep’ (N is a moraic nasal).

was impossible to locate the boundary between the two moras in the long vowel, it was aligned somewhere slightly after the middle, regardless of the vowel type and across the subjects.

In the items of CVV (‘VV’ of CVV is a diphthong), the F0 peak occurred within the vowel sequence, as in the CVR items, though its location appeared to depend on the vowels comprising the vowel sequence (see Figure 3.4). While the F0 peak occurred after the middle of the vowel sequence in almost all cases, it tended to occur later when the first vowel of the vowel sequence was a low vowel. This tendency was observed across the subjects.

Visual inspection revealed that ososagari was realised differently according to the syllable/mora structure: CV+CV and CVCV (with ososagari) vs. CVN, CVR and CVV (without obvious ososagari). It was also (though impressionistically) observed that the F0 peak was aligned with certain places in the segmental sequence in fairly consistent manners depending on the syllable/mora structures: with the beginning of the vowel of the syllable following the accented syllable for CV+CV and CVCV; with the end of

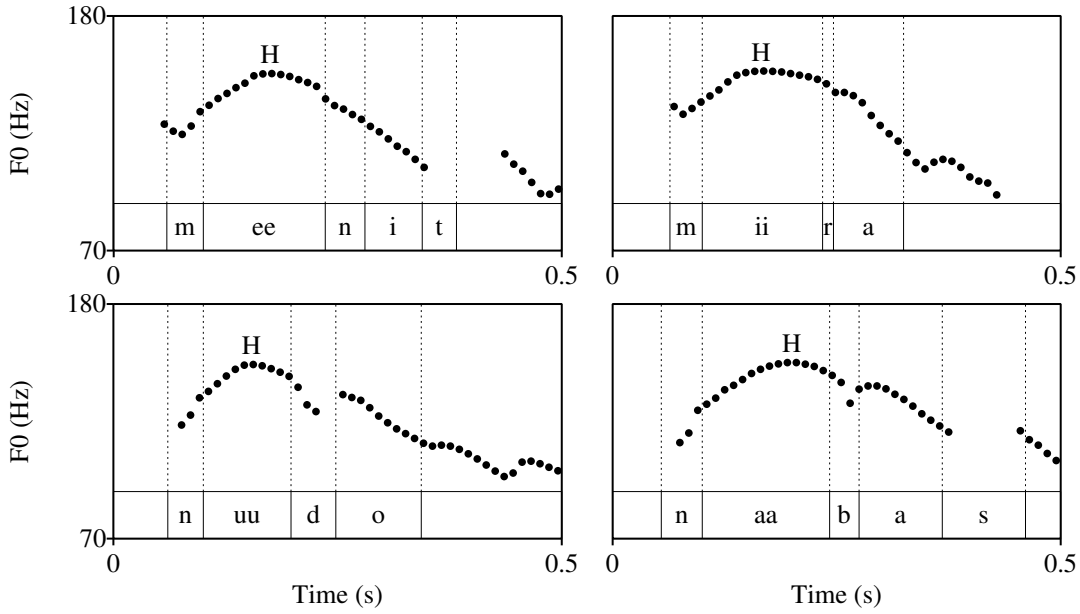


Figure 3.3: Examples of the F0 peak for initial-accented words which begin with the CVR sequence. 'R' is the second part of a long vowel. /meeniti/ 'the anniversary of somebody's death', /nuudoru/ 'noodle' /miira/ 'mummy' and /naabasu/ 'nervous'.

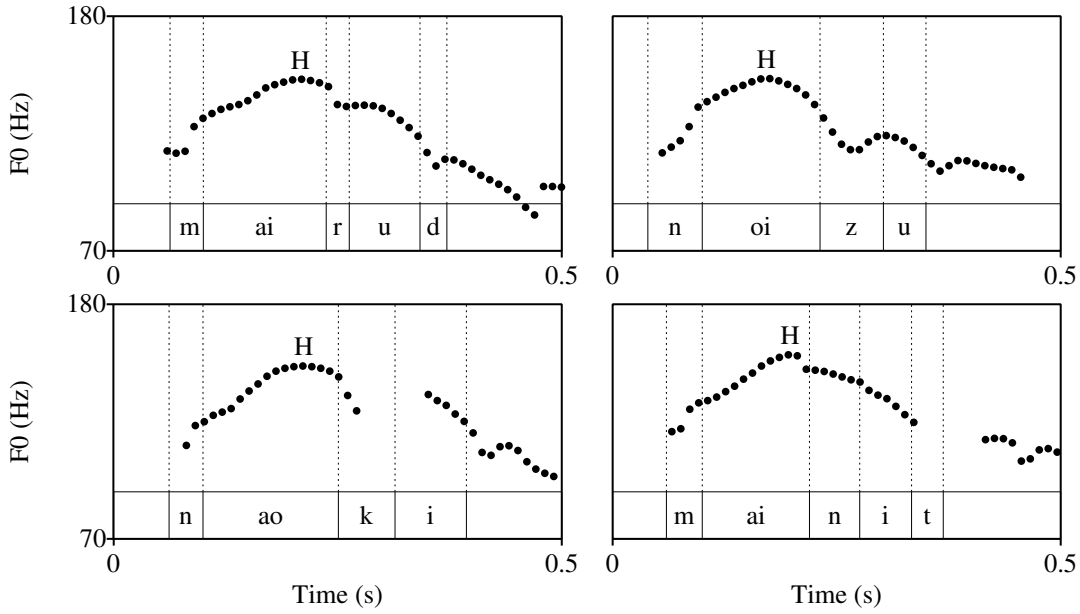


Figure 3.4: Examples of the F0 peak for initial-accented words which begin with the CVV sequence. 'VV' of CVV is a sequence of two different vowels. /mairudo/ 'mild', /nao+ki/ 'Naoki (male name)', /noizu/ 'noise' and /mai+niti/ 'daily'.

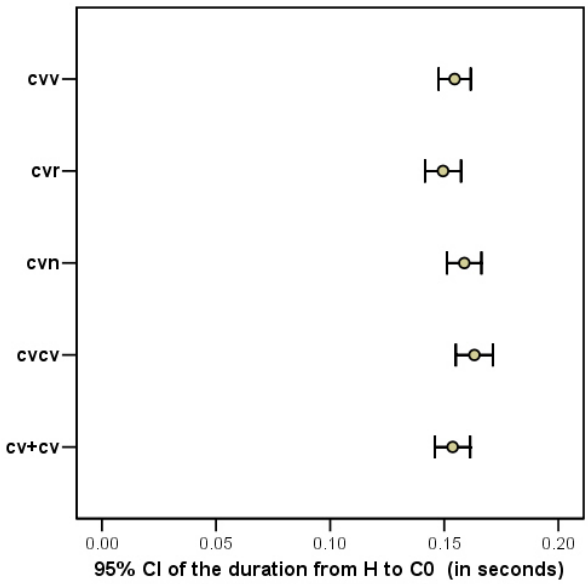


Figure 3.5: Mean duration from C0 to H in seconds. 'C0' is the beginning of the target word, as shown in Figure 2.2 of Chapter 2. The value zero in the graph amounts to C0.

the vowel of the accented syllable for CVN; in the middle of the vowel of the accented syllable for CVR and CVV.

3.2.2 Alignment of the F0 maximum

Figure 3.5 shows the alignment of H to C0.[6] The F0 peak was on average aligned 152 ms after C0 in the CV+CV; 162 ms after C0 in the CVCV cases; 157 ms after C0 in the CVN cases; 148 ms after C0 in the CVR cases; and 155 ms after C0 in the CVV cases. The data were analyzed in a two-way (5×7) mixed design ANOVA, with items as the random factor, Structure (structures of the target syllable) as a within-items factor, and Speaker as a between-items factor.[7] The ANOVA showed that there was no significant effect of Structure: $F(4, 216) = 2.409$; $p = 0.05$: no significant interaction between Structure and Speaker: $F(24, 216) = .883$; $p = .626$. The mean values for Speaker differed significantly beyond the 1% level: $F(6, 54) = 9.462$; $p < 0.0005$.[8]

[6]The tables of the measurements described in the current chapter are all in Appendix A.

[7]Pseudo F-test is an alternative test to the ANOVA used here, but the ANOVA was employed because of its accessibility in SPSS 12.0.

[8]Post hoc comparisons using the Tukey test indicated that the mean for RM was significantly different from that of all the speakers except TN, or vice versa. The mean for TN was significantly different from that of AK, FS, NI and ST, or vice versa.

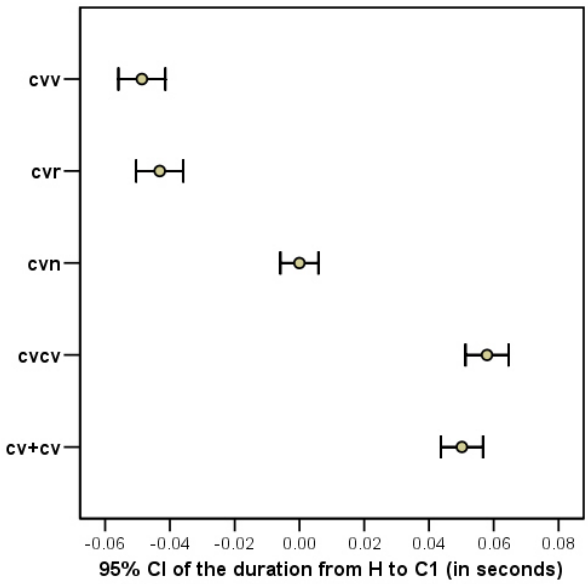


Figure 3.6: Mean duration from H to C1 in seconds. 'C1' is the end of the initial mora (as well as the initial syllable) for CV+CV and CVCV; it is the end of the initial mora (not the initial syllable) for CVN; it is the end of the initial syllable (not the initial mora) for CVR and CVV, as shown in Figure 2.2 of Chapter 2. 'cv1' stands for CV+CV, and 'cv2' for CVCV.

Figure 3.6 shows the alignment of H to C1. The F0 peak was on average aligned 47 ms after C1 in the CV+CV cases; 57 ms after C1 in the CVCV cases; just about C1 (0.2 ms before C1) in the CVN cases; 53 ms before C1 in the CVR cases; and 48 ms before C1 in the CVV cases. Again, the data were analyzed in a two-way (5×7) mixed design ANOVA, with items as the random factor, Structure (structures of the target syllable) as a within-items factor, and Speaker as a between-items factor. The ANOVA showed that the mean values for Structure differed significantly beyond the 1% level: $F(4, 208) = 323.498; p < 0.0005$. There was no significant interaction between Structure and Speaker: $F(24, 208) = .731; p = .816$. The mean values for Speaker also differed significantly beyond the 1% level: $F(6, 52) = 15.896; p < 0.0005$. A post hoc Bonferroni test revealed a significant difference between the target sequences ($p < 0.05$). The five target sequences thus can be grouped into three types—i.e. before, around, and after C1—in terms of the F0 peak alignment.

Figure 3.7 shows the alignment of H to V1. The F0 peak was aligned 5 ms before V1 in the CV+CV cases; 1 ms before V1 in the CVCV cases; 100 ms before V1 in the CVN cases; 92 ms before V1 in the CVR cases; and 95 ms before V1 in the CVV cases.

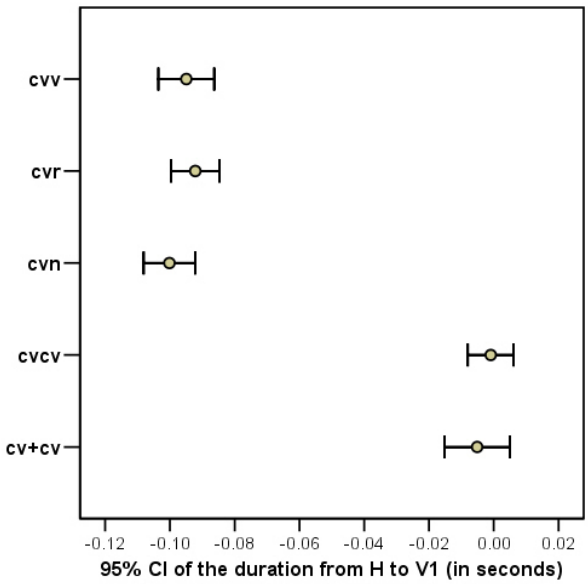


Figure 3.7: Mean duration from V1 to H in seconds. 'V1' is the beginning of the vowel of the second syllable across the groups, as shown in Figure 2.2 of Chapter 2. The value zero in the graph amounts to V1. 'cv1' stands for CV+CV, and 'cv2' for CVCV.

Again, the data were analyzed in a two-way (5×7) mixed design ANOVA, with items as the random factor, Structure (structures of the target syllable) as a within-items factor, and Speaker as a between-items factor. The ANOVA showed that the mean values for Structure differed significantly beyond the 1% level: $F(4, 196) = 200.842; p < 0.0005$. There was no significant interaction between Structure and Speaker: $F(24, 196) = .812; p = .719$. The mean values for Speaker also differed significantly beyond the 1% level: $F(6, 49) = 13.024; p < 0.0005$. A post hoc Bonferroni test revealed a significant difference between the target sequences ($p < 0.05$), and it can be regarded that there are two types of the F0 peak alignment: one around V1 for CV+CV and CVCV, the other at approximately 100 ms before V1 for CVN, CVR and CVV.

The F0 peak of CVR and CVV occurred in the two-mora vowel of the accented syllable. Figure 3.8 shows the proportional F0 peak location within the vowel. The F0 peak for both CVR and CVV occurred at a point about 70 % through the vowel. To see if the F0 peak synchronised with the mora boundary in the diphthongs (i.e. CVV), the F2 transition of the diphthongs was examined.

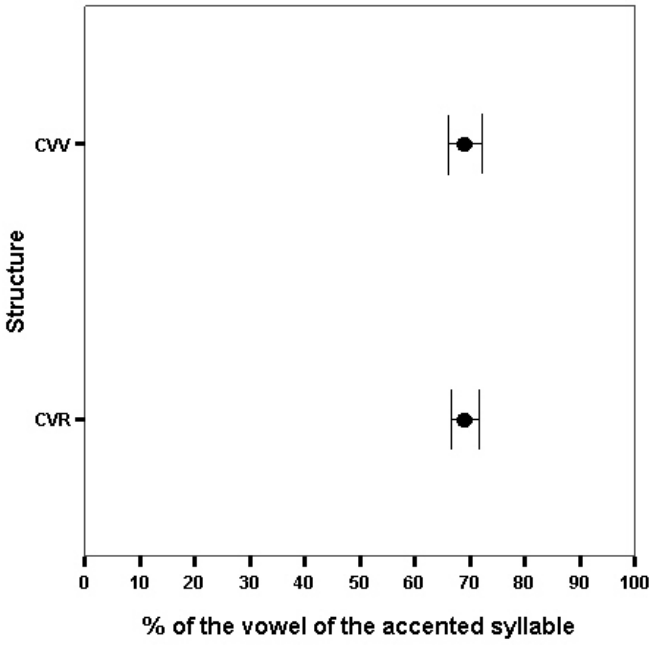


Figure 3.8: Proportional F0 peak location within the vowel (V0 to C1 in Figure 2.2) of the accented syllable for the CVR and CVV sequences. The horizontal scale shows the proportional vowel duration.

The data were analysed with a paired-samples *t* test; proportional locations of the most prominent F2 transition point[9] and the accentual F0 peak were compared. The most prominent F2 transition ($M = 0.60, SD = 0.09$) occurred earlier than the accentual F0 peak ($M = 0.69, SD = 0.16$). The paired-samples *t* test showed significance beyond the 1% level: $t(64) = 5.366, p < 0.0005$. Although item-by-item visual inspection revealed that the accentual F0 peak occurred at around the same point as the most prominent F2 transition of the diphthong, the test showed that the F0 peak was aligned with a point slightly later than the mora boundary in the diphthong.

One of the important findings of the data above is that all the speakers showed similar alignment patterns across all the measurements (the alignment of H with C0, C1, and V1) and across all the syllable/mora structures of the target accented syllable. First of all, as the ANOVAs demonstrated above, there were no interactions between Structure and Speaker in any of the three measurements. Together with the result that there were

[9]The most prominent F2 transition point was manually labelled for each item.

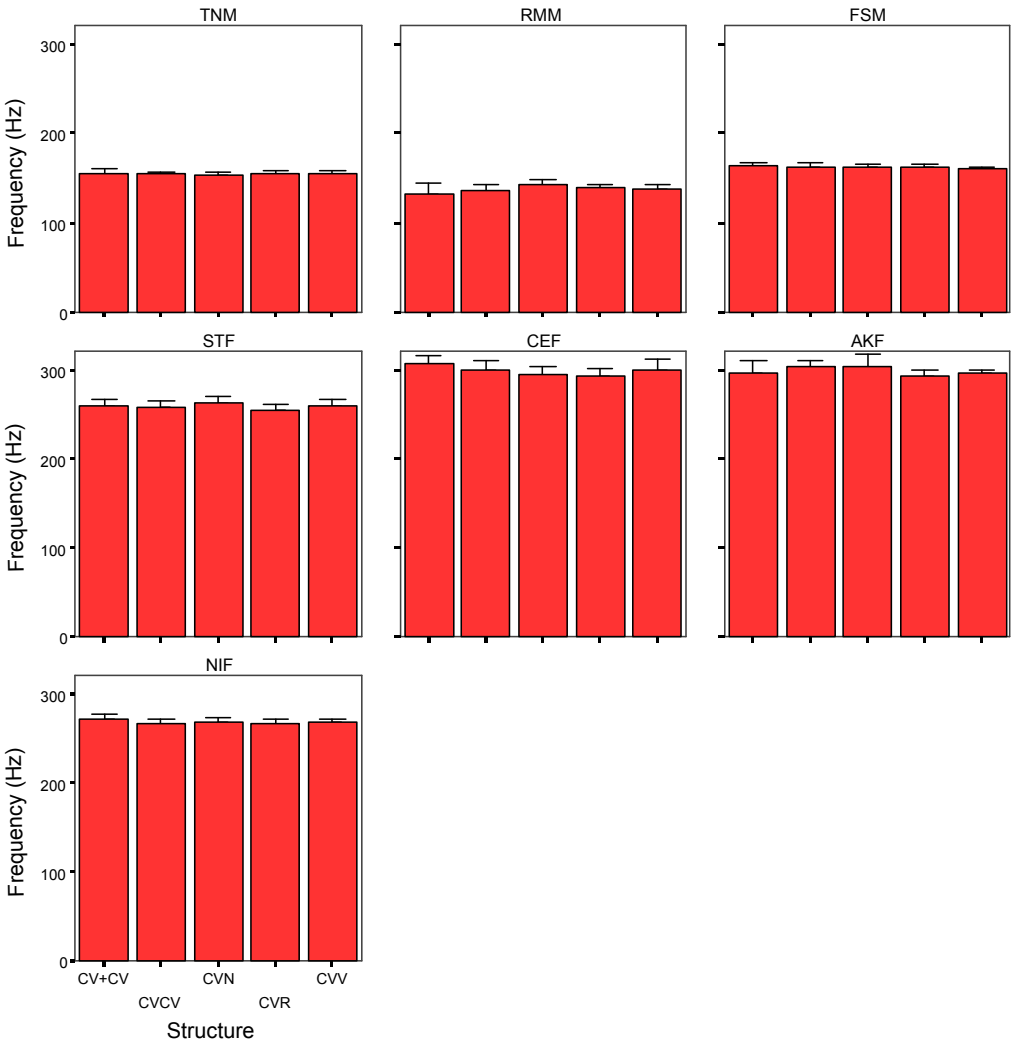


Figure 3.9: Mean F0 peak values in Hz. Each panel shows the data of individual speakers. The last letter (‘M’ or ’F’) of each speaker indicates gender (male or female).

no interactions between Structure and Speaker, these observed general patterns show that the F0 peak alignment was due to the syllable/mora structure of the accented syllable across all the speakers (I will return to these overall patterns in the discussions below about segmental anchoring and the appropriate measurement).

As supplementary information, Figure 3.9 shows the mean F0 peak values. The F0 value in Hz at H was used as a measurement, rather than the F0 change from the F0 valley to the F0 peak, since it was difficult to locate the F0 valley for L in some of the data (see

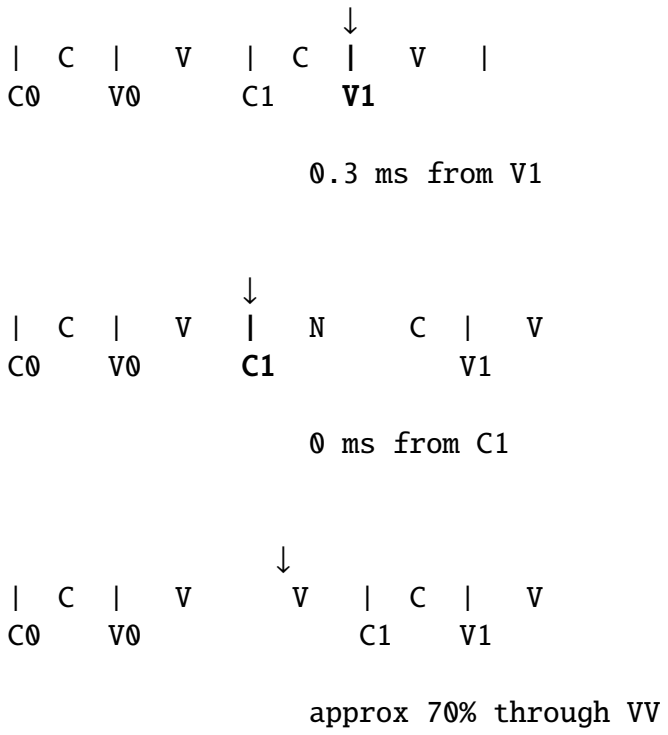


Figure 3.10: Schematic representation of alignment points (as shown with a downward arrow). The first sequence is for CV+CV and CVCV; the second for CVN; the third for CVR (a long vowel) and CVV (a diphthong). The segmental measurement points for the first two sequences are shown in bold. The F0 peak location for the third sequence is the ratio of H, to the period from V0 to C1 (the vowel duration). The values below are the average for each group.

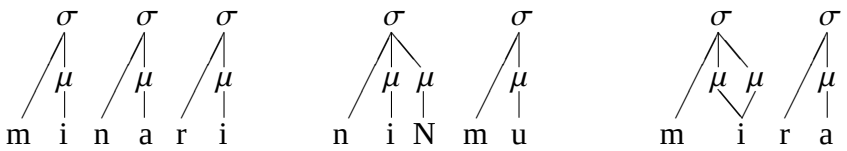
Section 3.2.3 for details). The data were analyzed by individual speakers, using one-way ANOVAs, with items as the random factor and Structure as a single within-items factor. There was no significant difference between the structures for all the speakers except one (Speaker FS: $F(4, 20) = 3.648; p < 0.05$). Compared to the data of the other speakers, it seems that the significant effect found in the data of Speaker FS is negligible. Apart from the individual variation in F0, the peak value can be regarded as being consistent regardless of the syllable/mora structures.

The above data on the F0 peak alignment can be interpreted in at least two different ways. Firstly, based on the data about the alignment of H to C1 and of H to V1, as well as the visual inspection, the F0 peak was consistently aligned with a specific point in the segmental string, depending on the syllable/mora structures of the accented syllable. Figure

3.10 shows these alignment patterns according to the syllable/mora structure schematically. As can be seen, for CV+CV and CVCV, V1 is the segmental landmark for the F0 peak; C1 for CVN; within the vowel of the accented syllable (about 70% of the vowel) for CVR and CVV. Moreover, in terms of the organisation of the mora and syllable in Tokyo Japanese, it seems feasible to interpret these alignment patterns as a consequence of segmental anchoring. According to one of the proposals of moraic phonology (Hayes 1989), the organisation of the mora and syllable is as follows.

(3.2)

a. b. c.



Given that onset consonants are attached directly to a syllable node, rather than to a mora, it can be claimed in terms of the current data that the F0 peak is anchored to the beginning of the second mora across all the syllable/mora sequences. This may lead to a more precise description of ososagari (i.e. peak delay in Japanese). That is, ososagari is one of the consequences of the anchoring of a pitch accent to the beginning of the second mora.

On the other hand, there is an alternative characterisation, because the mean duration from C0 to H was not significantly different between the syllable/mora structures. It is possible to claim that the F0 peak location resulted from some constant F0 rise duration, which made the F0 peak occur by accident at a specific point in the segmental sequence, and appear to show the consistent alignment described above. However, the ANOVA showed an almost significant effect of Structure ($p = .05$). Moreover, considering the possibility that this non-significance resulted from the greater variance due to the choice of a larger domain (C0 to H), as pointed out in Atterer and Ladd (2004), this non-significance is questionable, and therefore the claim of constant F0 rise duration is less convincing.

In the light of the results discussed above, although segmental anchoring seems more compelling so far, further analysis is necessary in order to decide which explanation is more plausible. Figure 3.11 shows the mean duration of the segments of the target sequences. The aligned points described above—i.e. V1 for CV+CV and CVCV, C1 for CVN, and 70% through the vowel for CVR and CVV—all came around 150 ms after C0,

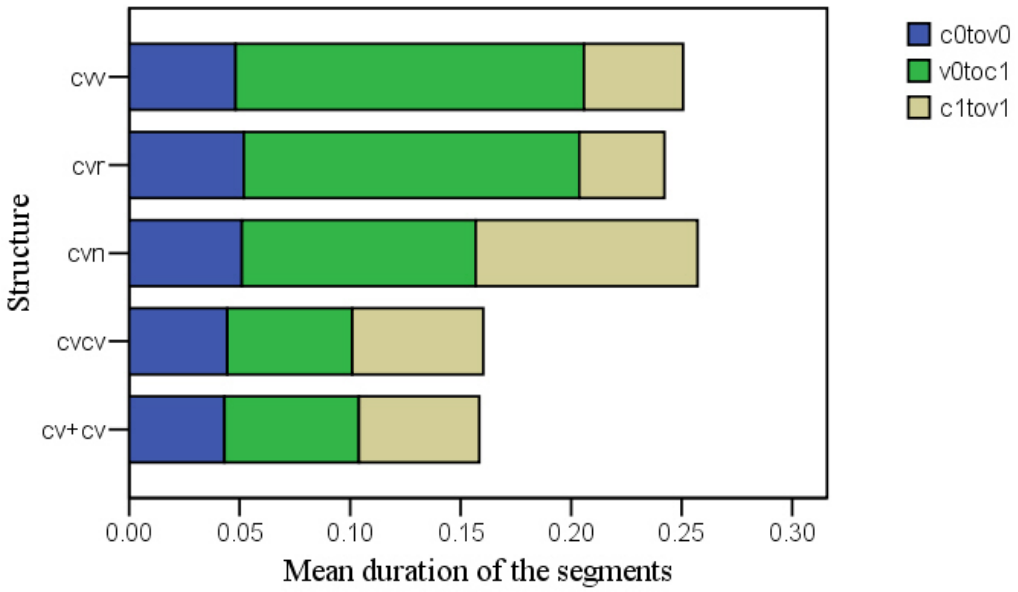


Figure 3.11: Mean duration of the segments of the target sequences. The value zero amounts to the beginning of the target sequence (i.e. C0). 'c0tov0', 'v0toc1' and 'c1tov1' are from C0 to V1, V1 to C1, and C1 to V1 in Figure 2.2 of Chapter 2, respectively.

as also demonstrated in the data on the alignment of H to C0. Taking this into account, it seems more plausible to say that the F0 peak location arose from constant F0 rise duration (and that the observed alignment was at most a coincidence).[10]

However, it is also very important to note that the vowel duration of the CVN sequence was considerably longer than that of the first syllable of the CV+CV and CVCV sequences. While the initial 'CV' of both CV+CV and CVCV, and CVN is one mora (though 'CV' of CV+CV and CVCV is also a syllable), the vowel duration of CVN is about twice as long as that of CV+CV and CVCV (also note the predictably longer vowel duration of the CVR and CVV sequences which consist of two moras). This vowel lengthening of CVN is reported in Homma (1981), and has been taken as evidence for a mora-based control of speech in Tokyo Japanese. While it is not clear why the accentual F0 peak was aligned with the end of this lengthened vowel of the CVN sequence regardless of the type of vowel, it might be possible to say that the F0 peak and the segmental landmark coordinated (or constrained) each other in order to achieve this alignment.

[10]It must be noted that constant F0 rise duration is only one of a possible set of non-alignment hypotheses, and there may be other non-alignment characteristics which explain the F0 patterns observed in the current data. One of the main reasons for the choice of constant F0 rise duration here is that it has been employed to express the pitch accent in some of the important studies (e.g. Fujisaki 1983).

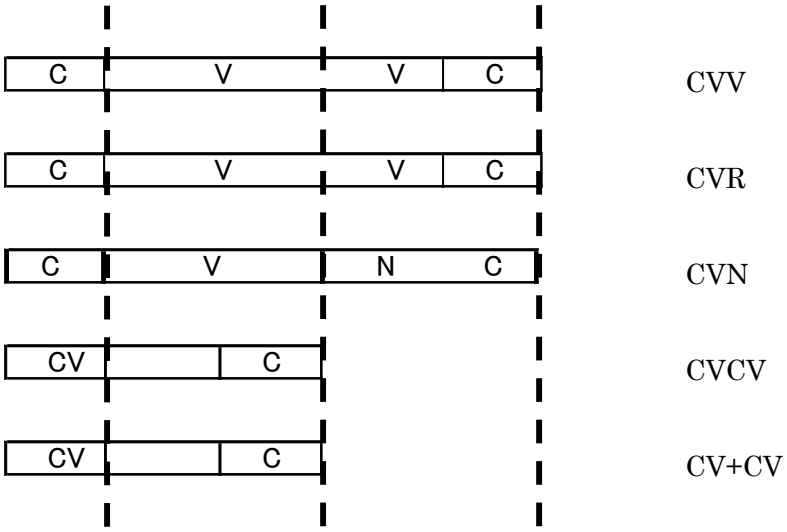


Figure 3.12: Schematic representation of segmental duration shown in Figure 3.11. 'V' (and 'N') phonologically amount to one mora on Hayes' terms. 'C' means an onset consonant.

The durational analysis of the segments of the target sequences here raises further issues of timing control in Tokyo Japanese. Looking at the pooled data shown in Figure 3.11, the duration from V0 to V1 for CV+CV and CVCV, the duration from V0 to C1 for CVN, and 70% of the duration from V0 to C1 for CVR and CVV seems quite similar. Moreover, though less convincingly, the duration from C1 to V1 and the duration of the last 30% of the vowel and the following consonant (C1 to V1) for CVR and CVV also seem similar. A graphical representation of Figure 3.11 is shown in 3.12. Based on these observations, it may be possible to suggest that the parts between the first and second dotted lines, and between the second and third, work as some sort of a timing unit. Moreover, these parts correspond to a mora (plus the onset consonant of the following syllable when there is) in the syllable/mora structure on Hayes' terms (see (3.2) above). That is, unlike the traditionally proposed timing unit (CV and the moraic phoneme), V and the moraic phoneme like N—plus the onset consonant of the following syllable, if there is one—are a timing unit, which tends to be of similar duration. To test this hypothesis, the durational data were analysed using a one-way repeated-measures ANOVA with items as the random factor, and Mora (four proposed domains marked by the vertical lines below) as a within-items factor.

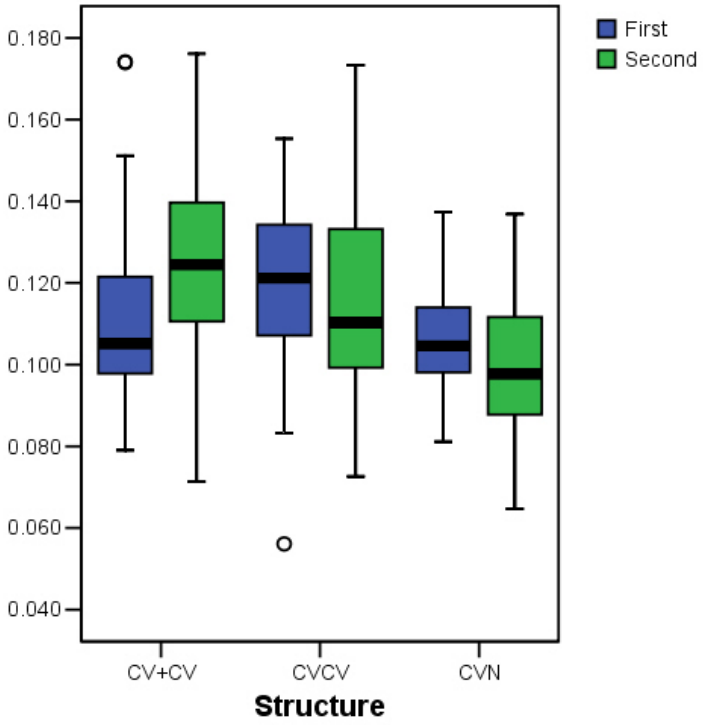


Figure 3.13: Boxplots of the duration (in seconds) of the proposed domains. 'First' means the first domain, and 'Second' means the second domain. See (4.1) for the detail.

(3.3)

```
#   C     V  .  C     V  .  C     V  .        -- CV+CV & CVCV
        |    1st    |    2nd    |

#   C     V     N  .  C     V  .              -- CVN
        | 1st |    2nd    |
```

The ANOVA showed that there was a statistically significant difference between duration of the proposed domains beyond the 1% level: $F(5, 295) = 14.2; p < 0.0005$, while a post hoc Bonferroni test revealed that the duration of first domain for CV+CV was not statistically significantly different from that of the first domains for CVCV and CVN. The data were also analysed using paired-samples *t* tests to compare between the first and second domains for CV+CV, CVCV and CVN separately. The *t* tests showed no significant difference between the first and second domains for CVCV: $t(61) = -0.508, p = 0.613$, and CVN: $t(69) = -1.950, p = 0.55$, but showed a statistically significant difference for

CV+CV: $t(66) = -5.144, p < 0.0005$. Figure 3.13 displays the data of the duration of the proposed domains. The results showed that the duration of the first domains were similar, and that the duration of the first and second domains were similar for at least CVCV and CVN. Although these are negative evidence, it may be interpreted that there were some sort of timing in action, and that as a consequence the vowel duration of the CVN sequence was made longer, and the F0 peak was aligned with the end of this timing unit, which made the F0 rise duration appear constant in the data.

3.2.3 Visual inspection of the F0 minimum

There were a number of cases which presented problems in carrying out visual inspection. First of all, the amount of the data of the F0 valley for L was half of that of the F0 peak for H, since it was not practical to locate the F0 valley for L reliably as an F0 turning point in utterance-initial items[11] (as described in Section 2.2.2 of Chapter 2, test words were put either at the beginning, or in the middle, of a test utterance).

Additionally, there was an unanticipated case in which it was impossible to track the target F0 valley. As described in Section of Chapter 2, speakers were only given minimal directions in the recordings in order to elicit natural renditions. Consequently, one speaker (Speaker CE) spontaneously put a brief pause between the target phrase and the phrase preceding it for most of the data, and another (Speaker ST) did this for about half of the data. This made it impossible to track the target F0 valley of the data, although there was no trouble tracking the target F0 peak. An example of this is shown on the left panel in Figure 3.14.

Since individual variations in the data were observed by the visual inspection, they were further explored with boxplots before statistical analyses in order to check the distributions of the F0 valley location. As shown in Figure 3.15, because Speaker FS clearly showed more idiosyncratic and much more variable alignment patterns, compared to the other speakers, his data was not included in the quantitative analyses. Due to these adverse cases described above, the data of four speakers (out of the seven) were eventually used.

It is also necessary to note that, even without a pause before an utterance-medial target phrase, there were items in which it was difficult to locate the F0 valley as a conspicuous F0 turning point. In these items, while the target F0 valley for L—as shown as the label 'L' on the right panel in Figure 3.14—was determined by a parabolic interpolation, there

[11] This can be seen in the examples in Figures 3.1, 3.2 and 3.3 of Section 3.2.1.

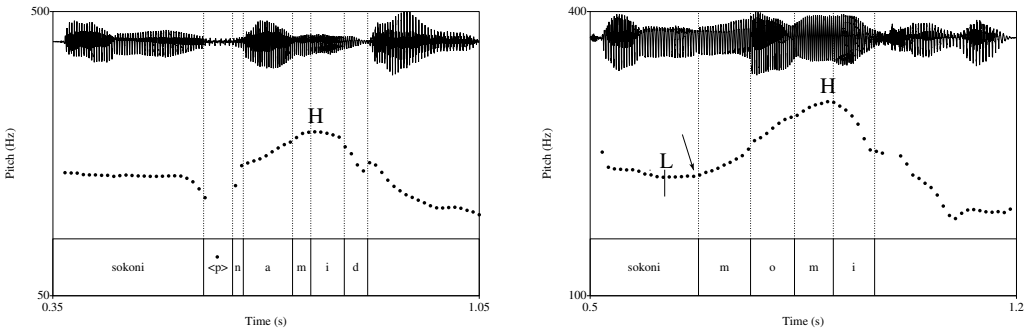


Figure 3.14: Examples of two kinds of cases difficult to locate the F0 valley. On the left is utterance-medial 'namida' from the data of Speaker CE, showing '<p>' (a brief pause) before the target phrase. On the right is utterance-medial 'momizi' from Speaker AK. 'L' is the point of the F0 valley for L (determined by a parabolic interpolation), while the point at the allow might visually (or subjectively) be an alternative F0 valley for L.

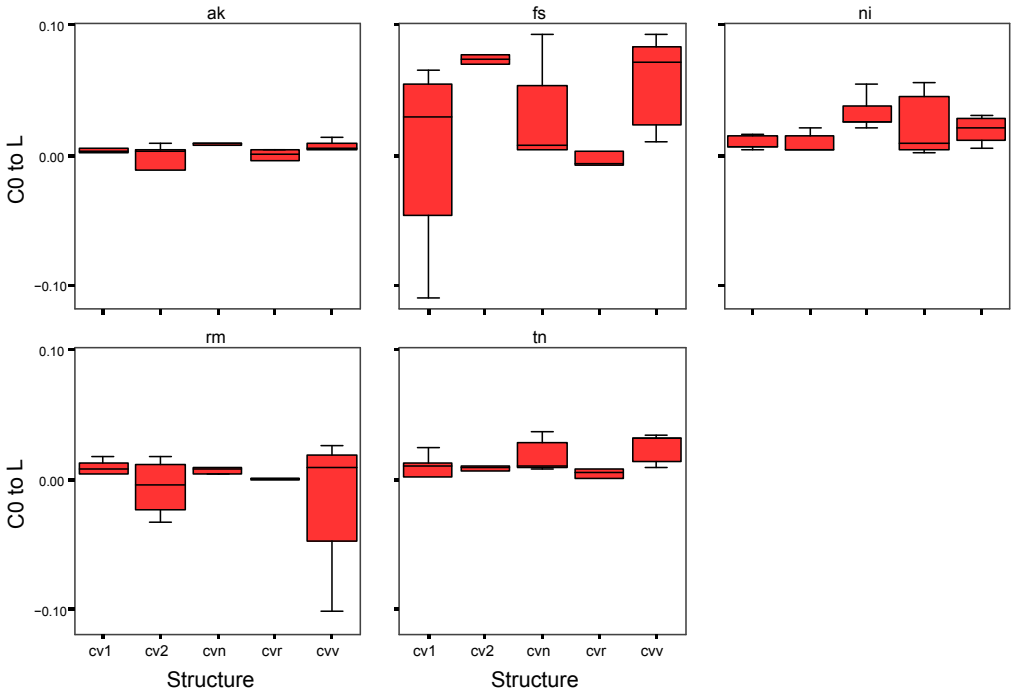


Figure 3.15: Boxplots of the alignment of L relative to C0. Each panel shows the data of individual speakers.

was an alternative point which seemed equally plausible for L in terms of the overall F0

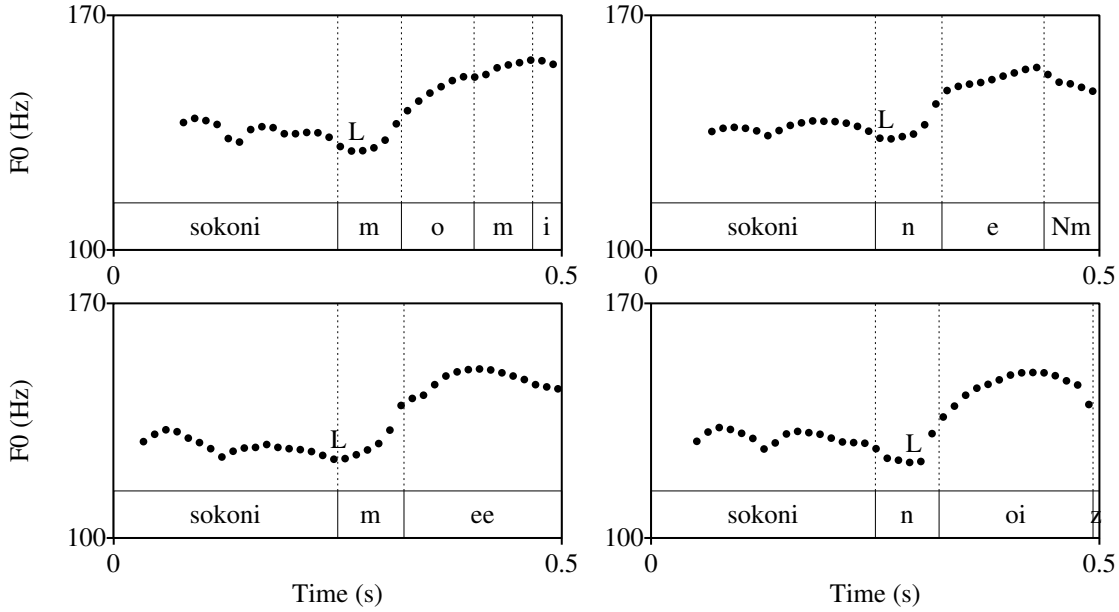


Figure 3.16: Examples of the utterance-medial F0 valley alignment: from the top on the left to the bottom on the right, /momizi/ 'red leaves' (CVCV), /neN+maku/ 'mucous membrane' (CVN), /mee+niti/ 'the anniversary of somebody's death' (CVR), and /noizu/ 'noise' (CVV). The first vertical dotted line indicates C0 (the beginning of the target sequence).

shape, as shown in the figure. The point identified by the parabolic interpolation was finally taken as the F0 valley for L for the objectivity in the measurement.

Figure 3.16 shows examples of the utterance-medial F0 valley alignment in the different syllable/mora structures of the accented syllable.[12] It was clear, based on the item by item visual inspection, that the F0 valley alignment was much more unstable than the F0 peak alignment. On the other hand, as shown in the figure, it seemed that the F0 valley occurred around C0 (the point at the first vertical dotted line) across the different syllable/mora structures. C0 was thus used as a measurement to examine the F0 valley alignment.

3.2.4 Alignment of the F0 minimum

Figure 3.17 shows the alignment of L to C0. The F0 valley was on average aligned 0.5 ms

[12] Again, for comparison, all the examples in Figure 3.16 are from the data of one male speaker (Speaker TN).

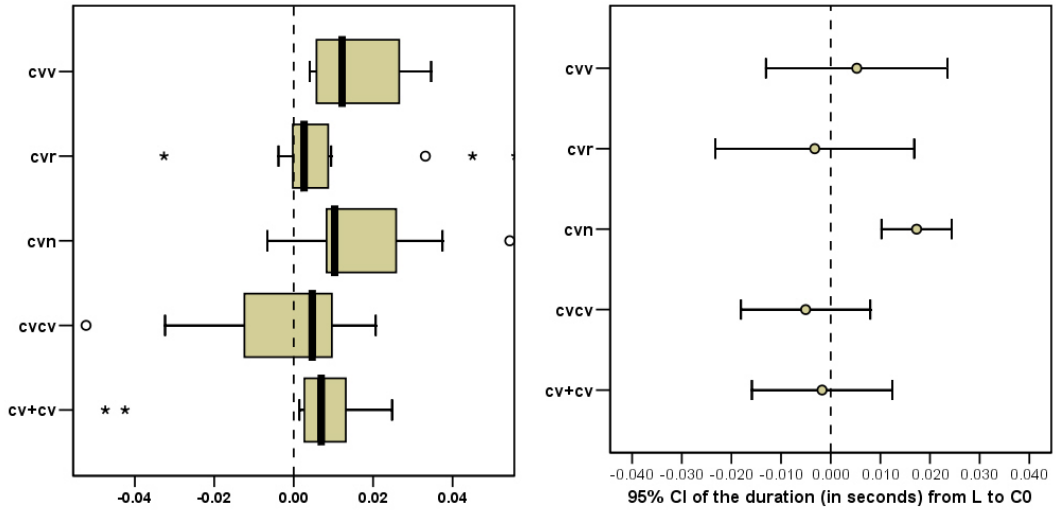


Figure 3.17: Alignment of L relative to C0. 'C0' is the beginning of the target accented syllable, as shown in Figure 2.2 of Chapter 2. The value zero in the graph corresponds to C0. The vertical scale shows the different structures, and the horizontal scale time (in seconds).

before C0 in the CV+CV; 4.0 ms before C0 in the CVCV cases; 17.7 ms after C0 in the CVN cases; 2.7 ms before C0 in the CVR cases; and 5.2 ms after C0 in the CVV cases (Overall 3.2ms after C0). The data were analyzed in a two-way (5×4) mixed design ANOVA, with items as the random factor, Structure (structures of the target syllable) as a within-items factor, and Speaker as a between-items factor. The ANOVA showed that there was no significant effect either both Structure: $F(4, 56) = 1.613; p = .184$, or Speaker: $F(1, 14) = .670; p = .427$. There was no significant interaction between Structure and Speaker: $F(12, 56) = .529; p = .887$. While the data showed some speaker idiosyncrasies (like Speaker FS) and measurement difficulties, the result showed the F0 valley was aligned just around C0. As can be seen in the figure, the F0 valley alignment for CVN appears different from that for the other structures, but there was no statistical difference.

The data of the minimum F0 level were also analyzed in a two-way (5×5) mixed design ANOVA, with items as the random factor, Structure (structures of the target syllable) as a within-items factor, and Speaker as a between-items factor. The ANOVA showed that there was no significant effect of Structure: $F(4, 56) = 1.613; p = .184$, no significant interaction between Structure and Speaker: $F(12, 56) = .529; p = .887$. As expected, there was a significant effect of Speaker: $F(4, 20) = 3.648; p < 0.05$). Apart from the

individual, particularly gender, variations in F0, it can be concluded that the minimum F0 level is stable regardless of the syllable/mora structures.

In sum, the results showed that the F0 valley is aligned with the beginning of the accented syllable, regardless of the syllable/mora structure of the accented syllable. The Low tone in this prosodic context is proposed to be associated with the edge of the accentual phrase, without the secondary association with the mora preceding the phrase boundary. Thus, the consistent F0 valley alignment at the beginning of the accented syllable seems to reflect the association of L to the edge of the accentual phrase in the prosodic structure.

3.3 Summary of the findings

- The levels of a boundary L and a pitch accent H are consistent regardless of the syllable/mora structures of the accented syllable, though there is individual variation.
- The F0 valley for a boundary L is consistently aligned with the beginning of the accented syllable—a few milliseconds after it on average—regardless of its segmental composition, while the data shows some individual variation.
- The F0 peak for a pitch accent H*+L is consistently aligned with a specific segmental landmark depending on the syllable/mora structures of the accented syllable. In particular the alignment patterns between the CV+CV, CVCV and CVN sequences are clearly distinct, regardless of the segmental differences in the initial syllable(s) (i.e. /m/ vs. /n/, and /i/, /e/, /a/, /o/ or /u/).
- The vowel duration of the accented syllable varies considerably depending on the syllable/mora structure of the accented syllable. The vowel duration of the accented syllable of .**CV**.CV. is about half of that of .**CVN**.CV., which is about two third of that of .**CVV**.CV. (the accented syllable is shown in bold type). Further durational analyses reveal that 'V.C' in .**CV**.CV., 'V' in .**CVN**.CV., 'NC' in .**CVN**.CV., the first 'V' in .**CVV**.CV. (which is the first mora of the two-mora vowel), and the 'V.C' in .**CVV**.CV. (which is the second mora of the two-mora vowel and the onset consonant of the following syllable) are all similar in duration.

Chapter 4

Tonal alignment in different speaking modes

4.1 Introduction

Changing ways of speaking, such as talking fast or raising the voice, is likely to affect the phonetic realisation of a pitch accent. For example, an F0 rise may become shorter in a situation where only limited time is available. On the other hand, as reviewed in Chapter 1, some of the previous studies have demonstrated the regularity of at least some aspects of a pitch accent regardless of change in speaking modes. Ladd *et al.* (1999), for example, showed invariant alignment of both the valley and peak of the prenuclear F0 rise with their segmental landmarks regardless of speech rate changes. Xu (1998) showed that, regardless of speech rate changes (and changes in segmental duration), the F0 peak of the rising tone in Mandarin Chinese is consistently aligned with a specific segmental landmark, and may shift earlier or later, synchronising with its segmental landmark. Knight (2002) found that the offset of the F0 plateaux (objectively defined section around the F0 peak) was consistently aligned with a specific segmental landmark regardless of pitch span expansion, while the F0 peak and the onset of plateaux were shifted later due to pitch span expansion.

The purpose of this chapter is to explore how tonal targets are influenced in different speaking modes, and whether there is any regularity regardless of change in speaking modes. The speaking modes of interest here are raised voice, fast speech rate, and local emphasis. As described in the previous chapter, the F0 peak for H in the normal spoken data is consistently aligned with a specific segmental landmark depending on the syllable/mora structure of the accented syllable. As shown in Figure 3.10 of Chapter 3, the alignment patterns can be divided into three types. The F0 peak is aligned with V1 (the beginning of the vowel of the syllable following the accented syllable) for CV+CV and

CVCV; with C1 for CVN (with the end of the first mora of the accented syllable); and about 70 percent into the two-mora vowel of the accented syllable for CVR and CVV. Based on comparisons between the alignment patterns found in the normal spoken data and those of the data with different speaking styles, the effects of the different speaking modes on the alignment of the tonal targets for a pitch accent, and phonetic invariance of tonal targets across different speaking modes are examined.

The data were collected under four different experimental conditions in order to elicit four types of data: normal, raised voice, fast speech rate and local emphasis on the target word. The same materials were used across the four experimental settings in order to make a direct comparison between the alignment patterns in normal speaking and those in the other speaking styles (sample words shown Table 3.1 of Chapter 3, and labelling scheme of target segmental sequences shown in Figure 2.2 of Chapter 2).[1]

4.2 Results

I first confirm in Section 4.2.1 that the manipulations were successful. Then I examine how the F0 peak alignment was influenced in the different speaking modes, based on the alignment pattens found in the normal spoken data.

4.2.1 Confirmation of speaking style manipulations

In order to confirm the rate manipulation, a comparison was made of the duration of the target syllable of the CV+CV and CVCV sequences (the duration from C0 to C1 of the first sequence in 2.2 of Chapter 2) between the normal and fast spoken data. A paired-samples *t* test was conducted to evaluate the effect of the two speaking styles, Normal and Fast, on the duration of the accented syllable. There was a statistically significant difference in the duration between Normal (M = 0.104, SD = 0.029) and Fast (M = 0.090, SD = 0.032): $t(117) = 3.117$, $p < 0.005$. This indicates that the speech rate manipulation was successful.

The mean F0 peak values of the normal spoken data were compared with those of the data with raised voice and local emphasis to verify their manipulations to confirm the pitch range manipulations. Figure 4.1 shows the mean F0 peak values between the speaking modes for each speaker. The data were analyzed using a two-way (3×7) mixed design ANOVA with Speaking Mode (Normal, Raised Voice and Local Emphasis) as

[1] See Chapter 2 for the full details of the material and experimental settings.

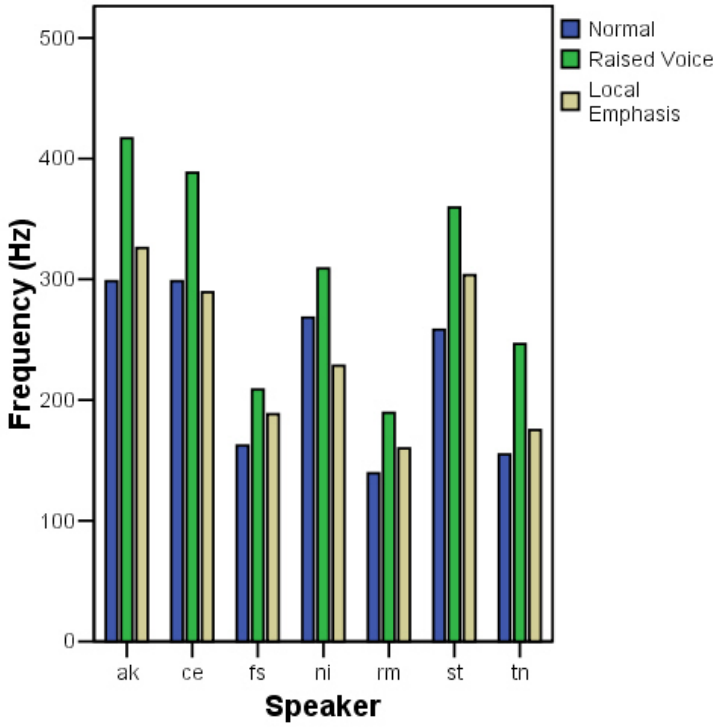


Figure 4.1: Mean F0 peak values in Hz between Normal, Raised Voice and Local Emphasis. 'fs', 'rm' and 'tn' are male speakers, and the rest are female speakers.

a within-items factor, and Speaker as a between-items factor. The ANOVA showed that the mean values for Speaking Mode differed significantly beyond the 1% level: $F(2, 634) = 3893.18$; $p < 0.0005$, and there was a significant interaction between Speaking Mode and Speaker beyond the 1% level: $F(12, 634) = 145.337$; $p < 0.0005$. The mean values for Speaker also differed significantly beyond the 1% level: $F(6, 317) = 4756.004$; $p < 0.0005$. A post hoc Bonferroni test revealed that the F0 peak values between the speaking modes were significantly different from each other ($p < 0.0005$). The ANOVA confirmed that all the speakers uttered in different manners depending on the speaking modes. As can be seen in the figure, all the speakers clearly raised their voice. On the other hand, in the data with local emphasis, while the mean F0 values for five speakers (AK, FS, RM, ST and TN) were higher than those of the normal spoken data, the values of the other two speakers were almost the same (Speakr CE) or lower (Speaker NI). Since there are various ways, including pitch range expansion, which Japanese speakers may take when producing an utterance with local emphasis, these two speakers are likely to have produced utterances in this experimental setting by means of some other way such as putting a pause before the target word or inserting an extra

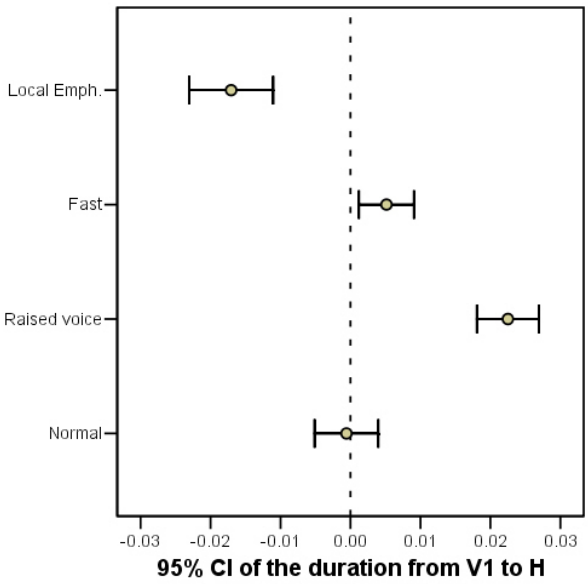


Figure 4.2: Alignment of H with V1 (in seconds) for CV+CV and CVCV across the different speaking modes. The vertical dotted line at the zero corresponds to V1 (the beginning of the vowel of the syllable following the accented syllable).

high tone at a certain place (this issue will be returned to later). Apart from these individual variations in the data with local emphasis, the result shows that the pitch range manipulations were successful.

4.2.2 Alignment of the F0 maximum

Alignment patterns were separately compared in terms of the syllable/mora structures to examine the effects of the different speaking modes. That is, rather than putting three factors (Structure, Speaking Mode and Speaker) into a three-way ANOVA, I conducted two-way ANOVAs, with the alignment of H with proposed segmental landmarks as a dependent variable, and Speaking Mode and Speaker as independent variables. The proposed landmarks were V1 for CV+CV and CVV; C1 for CVN; and C1 for CVR and CVV (because C1 was the closest to the F0 peak for CVR and CVV).

Alignment of H in CV+CV and CVCV

Figure 4.2 shows the alignment of H with V1 for CV+CV and CVCV across the different speaking modes. The F0 peak was on average aligned just about V1 (less than 1 ms

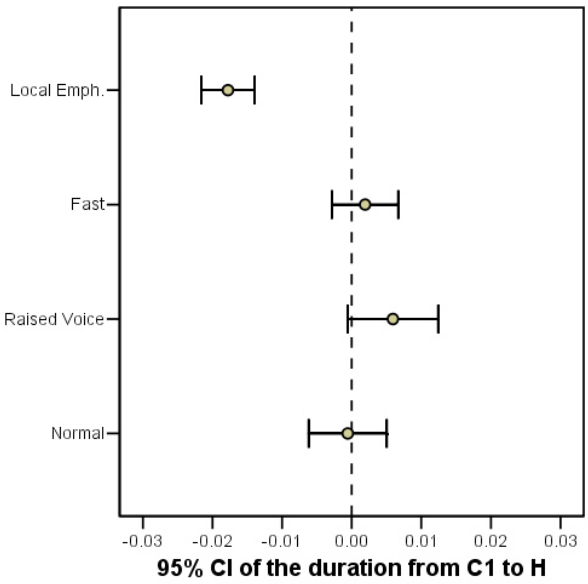


Figure 4.3: Alignment of H with C1 (in seconds) for CVN across the different speaking modes. The vertical dotted line at the zero corresponds to C1 (the end of the vowel of the CVN sequence).

before V1) in Normal; 23 ms after V1 in Raised Voice; 5 ms after V1 in Fast; 17 ms before V1 in the data with local emphasis.[2] The data were analyzed in a two-way (4×7) mixed design ANOVA, with items as the random factor, Speaking Mode as a within-items factor, and Speaker as a between-items factor. There was a significant effect for Speaking Mode: $F(3, 288) = 77.009; p < .0005$, and a significant interaction between Speaking Mode and Speaker: $F(18, 288) = 9.383; p < .0005$. There was also a significant effect for Speaker: $F(6, 96) = 4.273; p < .005$. Post hoc comparisons using the Bonferroni test revealed that the mean scores for the different speaking modes were significantly different from each other. The results showed that the F0 peak was shifted later by the effect of overall pitch raising. The increase of speech rate shifted the F0 peak slightly later, but the effect was not statistically significant. The F0 peak was shifted much earlier by the effect of local emphasis.

Alignment of H in CVN

Figure 4.3 shows the alignment of H with C1 for CVN across the different speaking modes. The F0 peak was on average aligned just about C1 (less than 1 ms before C1) in

[2]The tables of the measurements of the current chapter are all in Appendix B.

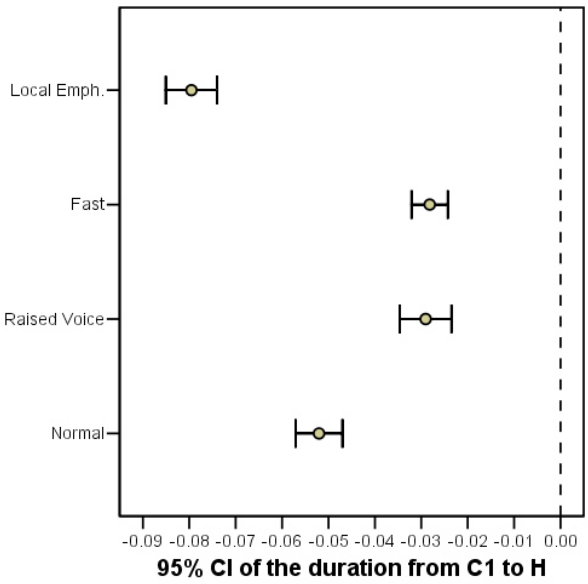


Figure 4.4: Alignment of H with C1 (in seconds) for CVR and CVV between the different speaking modes. The vertical dotted line at the zero corresponds to C1 (the end of the accented syllable) for CVR and CVV.

Normal; 6 ms after C1 in Raised Voice; 2 ms after C1 in Fast; 18 ms before C1 in Local Emphasis. The data were analyzed in a two-way (4×7) mixed design ANOVA, with items as the random factor, Speaking Mode as a within-items factor, and Speaker as a between-items factor. There was a significant effect for Speaking Mode: $F(3, 168) = 25.071$; $p < .0005$, and a significant interaction between Speaking Mode and Speaker: $F(18, 168) = 4.544$; $p < .0005$. There was also a significant effect for Speaker: $F(6, 56) = 11.893$; $p < .0005$. Post hoc comparisons using the Bonferroni test revealed that the mean score for Local Emphasis was significantly different from the other three modes, and not between the other three. The results thus showed that, compared to the normal spoken data, the F0 peak was shifted earlier by the effect of local emphasis. In the data of raised voice and fast rate speech, the peak was shifted only slightly later (6 ms and 2 ms, respectively), as can be seen in the figure, though their alignment was not significantly different from that of the normal spoken data. It can be noted that there was little effect of raised voice and fast rate on the F0 peak alignment in the CVN sequence.

Alignment of H in CVR and CVV

Figure 4.4 shows the alignment of H with C1 for CVR and CVV between the different speaking modes. The F0 peak was on average aligned 53 ms before C1 in Normal; 31 ms before C1 in Raised Voice; 28 ms before C1 in Fast; 79 ms before C1 in Local Emphasis. With reference to the mean value of the normal spoken data, the F0 peak occurred 23 ms later it in raised voice data; 25 ms later in fast speech data; 25 ms earlier in the data with local emphasis. The data again were analyzed in a two-way (4×7) mixed design ANOVA, with items as the random factor, Speaking Mode as a within-items factor, and Speaker as a between-items factor. There was a significant effect for Speaking Mode: $F(3, 336) = 131.889; p < .0005$, and a significant interaction between Speaking Mode and Speaker: $F(18, 336) = 6.365; p < .0005$. There was also a significant effect for Speaker: $F(6, 112) = 9.258; p < .0005$. Post hoc comparisons using the Bonferroni test showed that the mean scores for Raised Voice, Fast, Local Emphasis were significantly different from those for Normal, and that the mean scores for Raised Voice and Fast were not different from each other, as also seen in Figure 4.4. The results showed that the F0 peak was shifted later by the effect of raising the voice and speaking fast, while it was shifted earlier by the effect of local emphasis.

Relative F0 peak location between the speaking modes

The results above appear to show similar effects of the speaking modes on the F0 peak alignment between the syllable/mora structures. In order to combine these separate results into one comparable measure, relative F0 peak locations within a certain domain were calculated. Since it is impossible to establish a measurement common to all the syllable/mora structures within the first two moras, the ratio of the duration from V0 to H, to the duration from V0 to the vowel onset of the third mora was employed (shown below as the parts between the vertical lines).

(4.1)

```
#  C | V . C   V . C  | V  .        -- CV+CV & CVCV

#  C | V   N . C | V  .             -- CVN

#  C | V   V . C | V  .             -- CVR & CVV
```

This ratio expresses relative F0 locations between these portions. Figure 4.5 shows the distribution of the ratios between the speaking modes. As can be seen in the figure, the

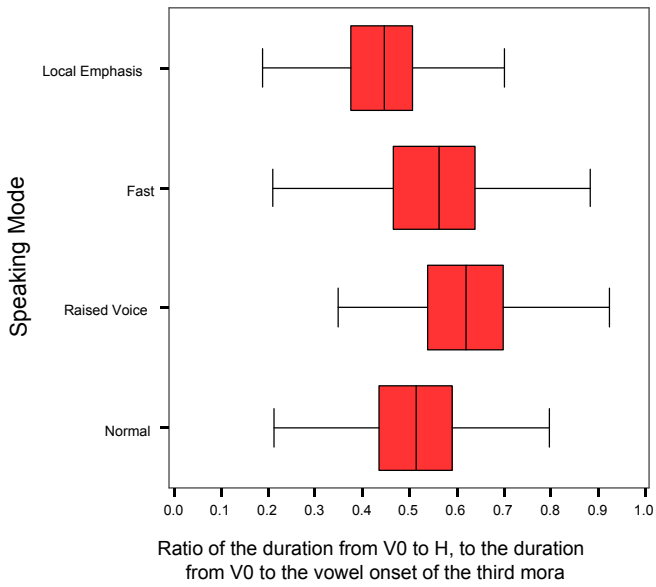


Figure 4.5: Ratio of the duration from V0 to H, to the duration from V0 to the vowel onset of the third mora

median ratio of Normal was at about 0.5, which indicates that the F0 occurred at the centre of this domain. Taken together with the result in Chapter 3 that the duration of the proposed timing units (V, or N, plus the onset consonant of the following syllable if there is one) were similar, this ratio of the normal spoken data also confirms that the F0 peak was aligned at the beginning of the second mora. That is, the F0 peak and the end of the timing unit proposed in Chapter 3 both took place at the middle of the domain in the normal spoken data.

Based on the alignment pattern for Normal, Figure 4.5 also enables us to examine the effects of the other speaking mode on the F0 peak alignment. The F0 peak occurred later in raised voice speech, compared to normal speech, while it occurred earlier in the data with local emphasis. The F0 peak occurred later in fast speech, but not as late in raised voice speech. All the alignment patterns between the speaking modes were consistent with those demonstrated in the separate results above.

Alignment of H with C0

significant effect of Speaking Mode: $F(3, 894) = 123.564$; $p < 0.0005$, and a significant interaction between Speaking Mode and Speaker: $F(18, 894) = 12.256$; $p < 0.0005$. The mean values for Speaker also differed significantly beyond the 1% level: $F(6, 298) = 31.062$; $p < 0.0005$. A post hoc Bonferroni test revealed that the mean values between the speaking modes were different from each other, except between Normal and Local emphasis ($p < 0.05$).

Figure 4.6 shows the mean duration from C0 to H between the speaking modes. All the panels show similar patterns: the duration from C0 to H for Raised Voice was longer, and for Fast was shorter, compared to that for Normal; the duration for Local Emphasis was similar to that for Normal. The data were analyzed using a two-way (4×5) within-subjects design ANOVA, with Speaking Mode and Structure as within-items factors. The ANOVA showed a significant effect of Speaking Mode: $F(3, 150) = 73.969$; $p < 0.0005$, and a significant effect of Structure: $F(4, 200) = 2.801$; $p < 0.05$, There was no significant interaction between Speaking Mode and Structure: $F(12, 600) = 1.414$; $p = 0.154$. A post hoc Bonferroni test revealed that the mean values for Normal and Local Emphasis were different from those for Raised Voice and Fast ($p < 0.05$), and that the mean values for Normal and Local Emphasis were not different from each other. Thus, these tests corroborated the patterns seen in the figure.

The normal spoken data of the alignment of H with C0, presented in Chapter 3, showed no statistically significant difference between the syllable/mora structures. As the alignment of H with C0 was regarded as being equivalent to the duration of the F0 rise in the previous chapter, this non-significance was considered to indirectly support constant F0 rise duration hypothesis. The present result clearly shows that the F0 rise duration was not constant between the speaking modes, which can be regarded against the constant F0 rise duration hypothesis.

Moreover, these alignment differences between the speaking modes seem to have resulted from changes in segmental duration due to the speaking styles, which is consistent with the segmental anchoring proposed in Chapter 3. For example, the earlier alignment in fast speech is likely to have resulted from the shorter segmental duration between C0 and the segmental landmark (such as V1 for CV+CV and CVV, or C1 for CVN) in fast speech. In order to explore this potential relationship between segmental duration and the duration from C0 to H, the segmental duration was also analysed. The data are shown in Figure 4.7. It can be clearly seen that the segmental duration, particularly the vowel duration,

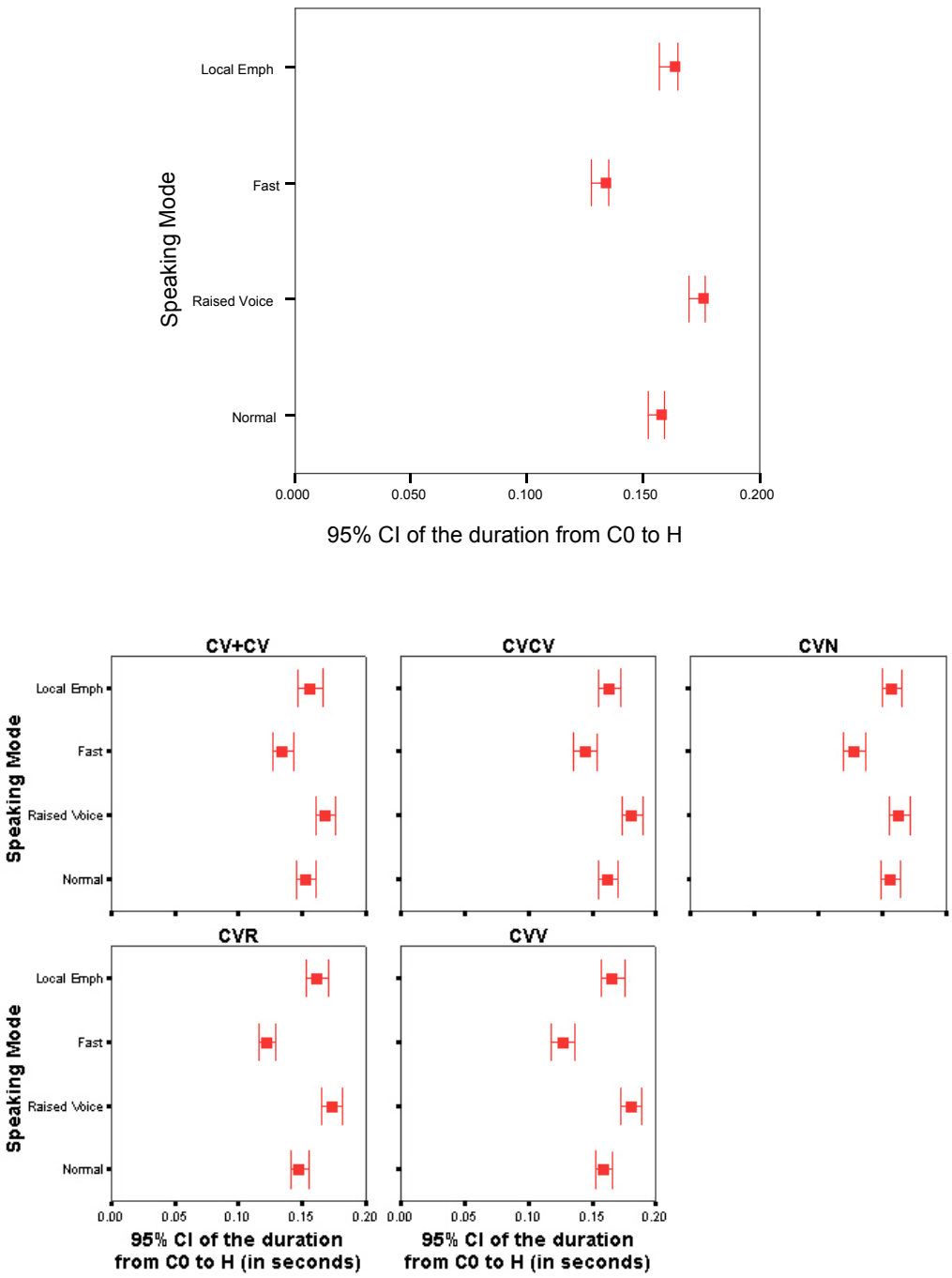


Figure 4.6: Alignment of H with C0 (in seconds). The large panel at the top shows the pooled data, and the small ones below it show the data in terms of the syllable/mora structures separately.

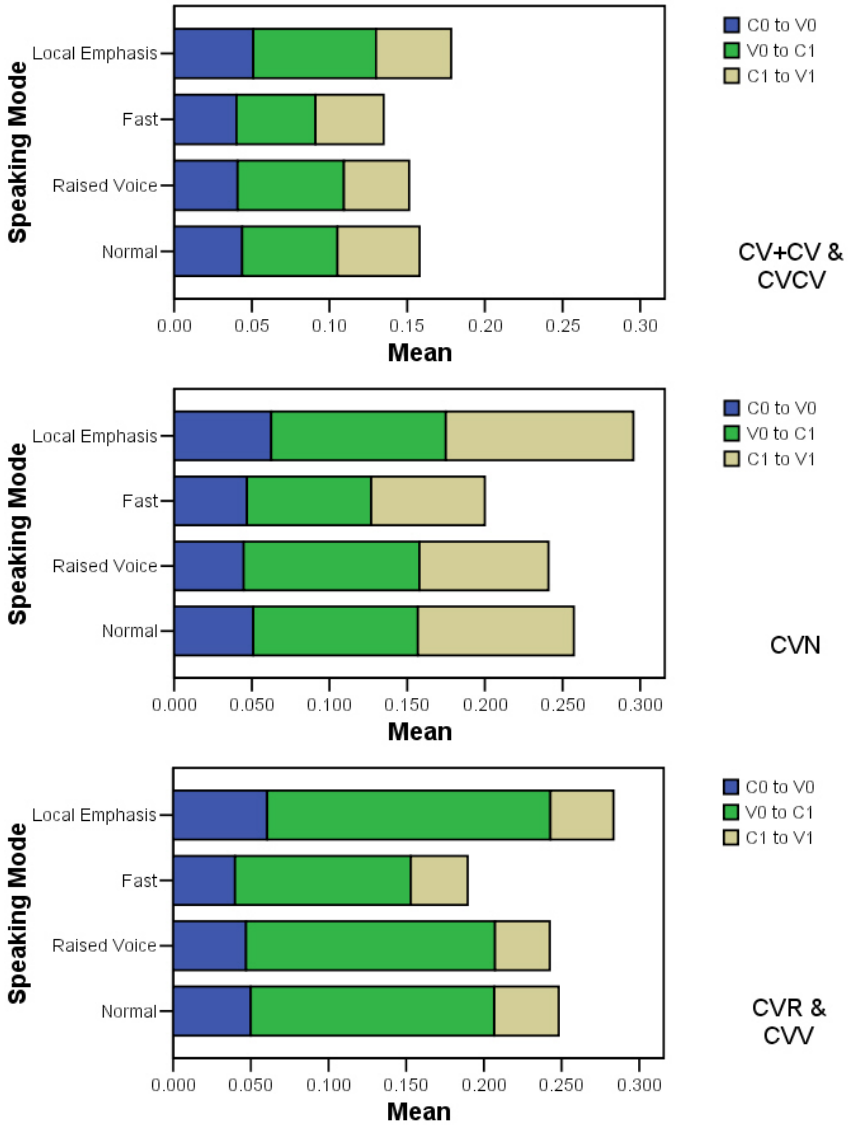


Figure 4.7: Segmental duration (in seconds) between the speaking modes.

was influenced by some of the speaking modes. In all the syllable/mora structures, the duration of the segments for Fast was considerably shorter, and the duration for Local Emphasis was considerably longer, compared to that for Normal. On the other hand, the segmental duration for Raised Voice was very similar to that for Normal.

Considering together the data of the alignment of H with C0 above, there are several points to note about the F0 peak alignment. For Raised Voice, since the duration from C0 to V1 for Normal and Raised Voice was similar and the duration from C0 to H for Raised Voice was longer than that for Normal, the F0 peak occurred later, compared to

Normal. For Fast, the duration from C0 to H and from C0 to V1 was shorter than that for Normal. On the other hand, as presented in the previous sections, the alignment of H with the segmental landmark was not different from that for Normal. These data of Raised Voice and Fast reaffirm the results of the F0 peak alignment presented above, and support the segmental anchoring to the proposed segmental landmarks.

The results for Local emphasis seem more complicated than those for Raised Voice and Fast. The data of the alignmnet of H with the proposed segmental landmarks showed that the F0 peak was aligned earlier. As shown in Figure 4.6, the alignment of H with C0 for Normal and Local Emphasis were similar. The segmental duration for Local Emphasis was considerably longer than that for Normal, as shown in Figure 4.7. In sum, the F0 peak location for Local Emphasis was determined by earlier alignment and longer segmental duration.

Although the alignment in speech with local emphasis showed a different behaviour from that in speech with the other speaking styles, there are some reasons for this. As mentioned in Section 4.2.1, there are various ways (including pitch range expansion) which Japanese native speakers may take when producing utterances with local emphasis, and they use one or more of the ways when speaking with local emphasis. Apart from pitch range expansion, these include putting a pause before and/or after the emphasised word/phrase, inserting an extra high tone around the end of the accentual phrase with the emphasised word, elongating the word, and so on. Also, it seems that these features influenced the F0 peak alignment of the current data in complicated ways (shifting the F0 peak earlier by an extra high tone around the end of the target accentual phrase, and longer segmental duration by elongation of the target word), which results in the earlier F0 peak alignment in locally emphasised speech.

4.3 Summary of the findings

In general, the orderly alignment behaviour seen in the data of Chapter 3 remains intact across different speaking modes.

- There were basically no significant differences in alignment between the normal spoken data and the fast spoken data.
- The F0 peak generally occurred slightly later in loud speech than in normal speech. However, the differences were very small in spite of the effect of overall pitch

raising, and it was clear that the syllable/mora structure of the accented syllable played an important role in the alignment, as in normal and fast speech.

- In speech with local emphasis, the F0 peak was aligned much earlier than in normal speech. Because the speakers took different procedures when producing an utterance with local emphasis (e.g. putting a pause before and/or after the target word, or inserting an extra high tone around the end of the target word), more factors are involved affecting the alignment and durational patterns. However, it was still possible to identify the alignment patterns depending on the syllable/mora structure of accented syllable, comparable to those seen in the other speaking modes.

CHAPTER 5

Tonal alignment in different accent patterns

5.1 Introduction

It has been suggested that ososagari is less likely to occur in non-initial-accented words (e.g. Sugito 1982). Since previous studies have focused more on the F0 peak alignment in initial accented words where ososagari is more likely to occur, it is not clear whether there is any regularity of the F0 peak alignment in non-initial-accented words. On the other hand, while it is reported that there are cases where the F0 peak for the phrasal H- occurs after the end of the associated mora (ososagari) (e.g. Venditti 2005), again, there is no quantitative study on alignment regularity of the F0 peak in unaccented words. Thus, the alignment of the F0 peak in structures with unaccented words and non-initial-accented words is largely unknown.

The aim of this chapter is to explore how the F0 peak is aligned with the segmental string in tonal structures other than those with the initial-accented words—i.e. in tonal structures of unaccented words and of words with accent on the syllable other than the initial. The results reported in the previous two chapters, which are based on the data of initial-accented words, show that the F0 peak for the pitch accent is consistently aligned with a specific segmental landmark depending on the syllable/mora structure of the accented syllable, and the alignment patterns were largely unchanged regardless of the changes of the speaking modes. The F0 peak alignment in the data of non-initial-accented words and unaccented words is examined, and compared to that in the data of initial-accented words.

The tonal structures of test utterances in the data of this chapter were different from those in the data of the previous chapters (though the same carrier sentences were used in the

experiments).[1] Differences include the presence/absence of accent, and the location of the accented syllable. There are three types of target tonal structures of the test utterances. One is of an accentual phrase which begins with a word with accent on the second syllable. In the terminology of Venditti (2005), the tonal structure of the accentual phrase can be represented as:

(5.1)

```
utterance-initial          %L  H*+L ...
utterance-medial      ... L%   H*+L ...
```

where the pitch accent H*+L is associated with the second syllable of the test word, and %L and L% is associated with the left and right edges of the acccentual phrase, respectively.

Another type is of an accentual phrase which begins with a word with the accent on either the third or fourth syllable. The tonal structure of the accentual phrase is:

(5.2)

```
utterance-initial          %L  H- H*+L ...
utterance-medial      ... L%   H- H*+L ...
```

where the pitch accent H*+L is associated with the third or fourth syllable of the test word; H- with the second mora; and %L and L% with one of the edges of the acccentual phrase and (secondarily) with the initial mora of the test word. So the difference between words with accent on the third syllable and the fourth syllable is the presence/absence of one syllable between the syllable with which the phrasal H- is associated and the accented syllable.[2]

The third type is of an accentual phrase which begins with a lexically unaccented word:

(5.3)

```
utterance-initial        %L H- ...   or         %wL H- ...
utterance-medial    ... L%  H- ...   or   ... wL%   H- ...
```

[1] See Section 2.2.2 in Chapter 2 for the full details of the carrier sentences.

[2] The absence of the F0 peak for the phrasal H- in the accentual phrase with pitch accent on the first or second syllable is explained in Pierrehumbert and Beckman (1988). In accentual phrases with pitch accent on the initial syllable, association of the phrasal H- with the second mora is blocked by autosegmental constraints. In those with pitch accent on the second syllable, the phrasal H-, as well as the accent H, is attached to the second mora. Pierrehumbert and Beckman conclude that the phrasal H- in both the cases would be phonetically invisible. In Venditti (2005), the phrasal H- is labelled 'on all unaccented phrases, and on accented phrases only where the H- is distinguishable from the high of the lexical accent' (p. 180).

where the phrasal H- is associated with the second mora of the target word, and that the boundary %L and L% are associated with both the initial mora and the accentual phrase, while wL% and %wL are associated only with the accentual phrase (wL% and %wL occur with words with a heavy initial syllable). Besides, all of them have a LH sequence at the beginning of the phrase.

Unaccented test words begin with one of the five different syllable/mora structures, analogous to those in the initial-accented target words used in the previous chapters: CV+CV, CVCV, CVN, CVR and CVV. On the other hand, non initial-accented test words only have a light accented syllable (e.g. /na.**ma**.ni.ku/ and /ma.me.**mo**.na.ka/), because it was impossible to find enough actually occurring words with various syllable/mora structures.[3]

In the following section, I first present the data of the accentual phrase (Section 5.2.1), and then the data on the alignment of the unaccented accentual phrase (Section 5.2.2). Note that, because the results of Chapters 3 and 4 showed that the alignment patterns between the speakers were similar and also because there was not enough time to carry out an analysis of the entire data, only the data of four speakers (two males and two females) were used.

5.2 Results

5.2.1 Accented accentual phrase

Visual inspection

As in Chapter 3, I began with a visual inspection of the alignment of the F0 targets in order to see how the F0 peak was aligned with the segmental string in different structures.[4] Figure 5.1 shows examples of the F0 peak alignment for the pitch accent on either the second, third or fourth syllable of the test word. For words with accent on the second syllable (the two panels on the top in the figure), it can be seen that the F0 peak was aligned around the end of the accented syllable. For words with accent on the third and fourth syllables (the examples on the middle and bottom panels in the figure), the F0 peak was aligned with the beginning of the vowel of the accented syllable, or somewhere in the middle of it. On the other hand, relatively large variation in the F0 peak alignment was informally observed in the data of non-initial-accented words, compared to those of

[3] See Section 2.2.1 in Chapter 2 for full details of the material.

[4] For comparison, all the examples in this section are from the data of one male speaker (Speaker TN).

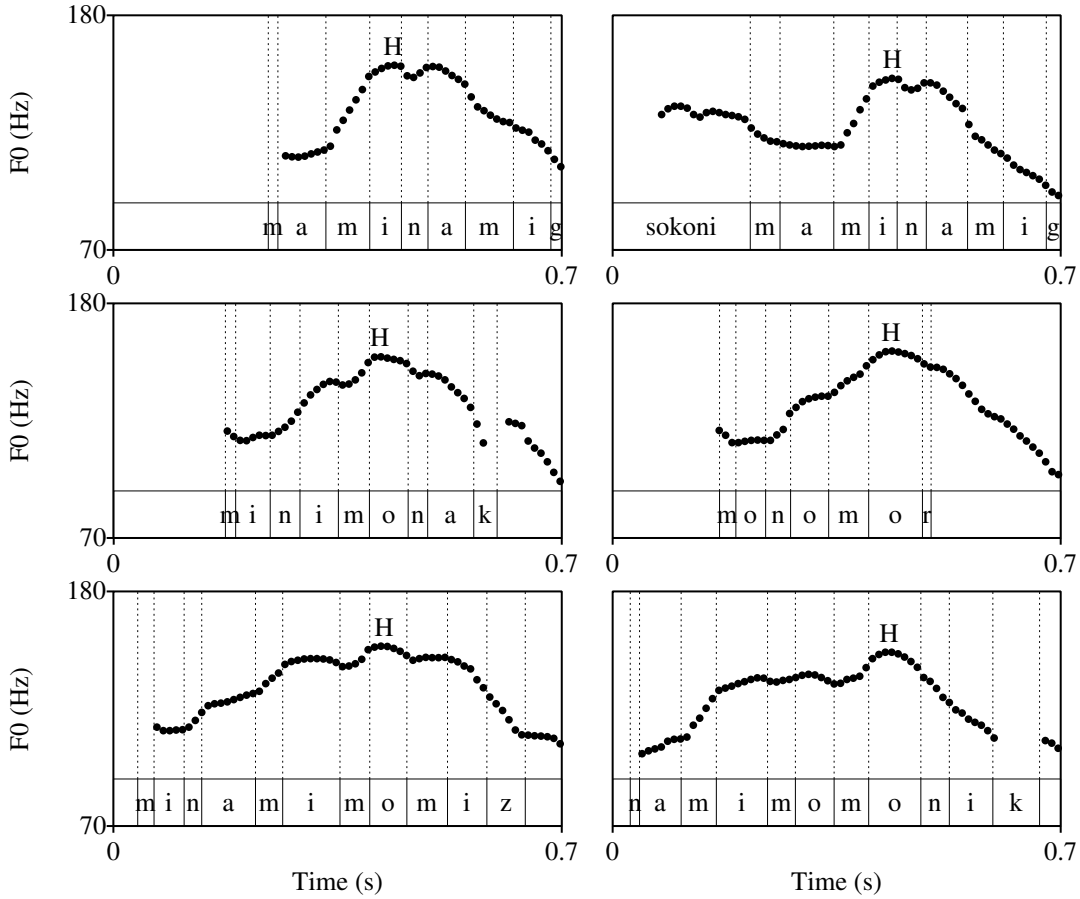


Figure 5.1: Examples of the F0 peak location for words with accent on the second (top), third (middle) and fourth syllable (bottom). /ma+minami/ ‘due south’ (utterance-initial), /ma+minami/ ‘due south’ (utterance-medial, /mini+monaka/ ‘mini bean-jam-filled wafers’, /monomorai/ ‘stye’, /minami+momizi/ ‘southern maple’, and /nami+momo+niku/ ‘regular-grade chicken thighs’.

initial-accented words. Based on visual inspection, it seems plausible to choose between the beginning of the vowel of the accented syllable, and the end of the syllable, as a measurement point. Both were used. In the following sections, accent patterns of the test words are given abbreviated names for clarity: accent on the second syllable for +A2; accent on the third syllable for +A3; accent on the second syllable for +A4.

Alignment of the F0 maximum for the pitch accent H+L*

Figure 5.2 shows the alignment of H relative to the end of the accented syllable. The F0 peak was on average aligned 24 ms after the end of the accented syllable for +A2; 35 ms

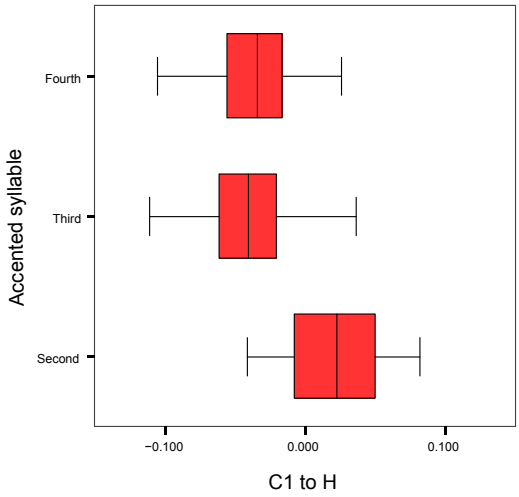


Figure 5.2: Boxplots of the alignment of H with the end of the accented syllable (in seconds). The value zero in the graph corresponds to the end of the accented syllable.

before the end of the accented syllable for +A3; 37 ms before the end of the accented syllable for +A4. Because there were quite a few missing cases for some speakers, particularly in the data of the +A2 words (some of the +A2 words have the alternative accent pattern (unaccented) and were arbitrarily pronounced with that pattern by these speakers), the data were analyzed in a one-way within-subjects ANOVA, with Accented Syllable as the within-items factor.[5] The ANOVA showed that there was a significant effect of Structure: $F(2, 42) = 23.230; p < 0.0005$. A post hoc Bonferroni test revealed that the mean values between +A2, and +A3 and +A4, were significantly different ($p < 0.05$), while there was no significant difference between +A3 and +A4.

The data was also examined in terms of the alignment of H with the beginning of the vowel of the accented syllable (see Figure 5.3). The F0 peak was on average aligned 87 ms after the end of the accented syllable for +A2; 45 ms after the beginning of the vowel of the accented syllable for +A3; 29 ms after the beginning of the vowel of the accented syllable for +A4. The data were again analyzed in a one-way within-subjects ANOVA, with Accented Syllable as the within-items factor. The ANOVA showed that there was a significant effect of Structure: $F(2, 42) = 11.668; p < 0.0005$. A post hoc Bonferroni test revealed that the mean values between +A2 and +A4 were significantly

[5]The tables of the measurements of the current chapter are all in Appendix C.

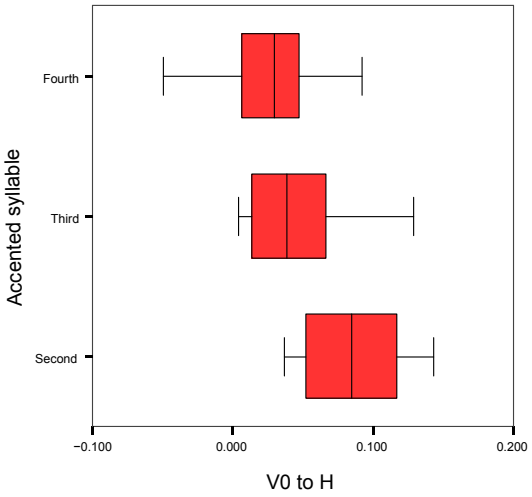


Figure 5.3: Boxplots of the alignment of H with the beginning of the vowel of the accented syllable (in seconds). The value zero in the graph corresponds to the beginning of the vowel of the accented syllable.

different ($p < 0.05$), while there was no significant difference between +A2 and +A3, and between +A3 and +A4.

The results above, based on the two measurements, clearly showed that the F0 peak alignment for +A2 was different from that for +A3 and +A4. On the other hand, it was not clear whether the F0 peak alignment between +A3 and +A4 was different. While one of the statistical tests demonstrated the difference, the other did not, and they do not appear different in Figures 5.2 and 5.3. To examine this further, the data of proportional peak locations in the vowel of the accented syllable between +A3 and +A4 was explored in a paired-samples t test. There was no significant difference between +A3 ($M = 0.51, SD = 0.49$) and +A4 ($M = 0.43, SD = 0.42$): $t(34) = 0.823, p = 0.416$. Hence, while there was a rather large variation of values, the F0 peak proportionally occurred around the centre of the vowel of the accented syllable for both +A3 and +A4.

Although there seems to have been no statistically significant difference in the F0 peak alignment between +A3 and +A4, it seems that the F0 peak for Fourth occurred earlier than that for +A3, which occurred earlier than that for +A2. This is not strictly consistent with what was found in the data of the initial-accented words discussed in Chapters 3 and 4 in which the accentual F0 peak was consistently aligned with the beginning of the

vowel of the syllable following the accented syllable in the sequence of light syllables like /**mo**.na.ka/.

This shifting of the F0 peak alignment may have correlated with the number of moras between the accented syllable and the end of the target word, given that there would have been some right-hand contextual effect. However, because all the target non-initial-accented words except one had two moras between the accented syllable and the end of the word, it was impossible to test this with the current data. It seems unlikely that some right-hand contextual effect caused the shifting the F0 peak earlier, since there were still alignment differences between these structures, as demonstrated above.

With the current data, it is hard to give conclusive evidence to these alignment differences between the non-initial-accented words. One possible explanation to this may be that, since tonal structures turned out to be different depending on the accent location (LHL where H was on the initial mora for the initial-accented word; LHL where H was on the second mora for the word with accent on the second syllable; LHHL where the first H was on the second mora, while the second H was on either third or fourth mora, for the word with accent on the third and fourth syllable), and thus the tonal sequence for +A3 and +A4 (and probably +A2) were looser, their F0 peak alignment was more variable and less stable than that for initial-accented words. Nevertheless, the present data was not sufficient to investigate this claim.

There are at least two interpretations on the alignment differences between words with accent on different syllables. One is that the alignment in initial-accented words does not result from segmental anchoring or the like. That is, they are just consequences of the interactions of factors affecting alignment. The other interpretation is that the alignment in the initial-accented words *does* result from segmental anchoring, but the alignment of the non-initial-accented words are influenced by some factors (and as a result the F0 peak is aligned earlier). In the light of the results of Chapters 3 and 4, the first interpretation is very unlikely. On the other hand, if the second interpretation is correct, that is, that the segmental anchoring in the initial-accented words is genuine, there are at least two questions to ask. One is what types of factors affect the alignment of non-initial-accented words. The other is why only the alignment of initial accented words is so consistent.

One possible answer to these questions is overall ‘tightness’ between the tonal events due to the differences of prosodic structure. For example, words with accent on the initial or second syllables have a tonal sequence, LHL, while words with accent on the third or fourth syllables have a LHHL sequence. So their accent patterns give rise to different

tonal sequences in an utterance. Moreover, there is no syllable between the phrasal H- and H*+L in words with accent on the third syllable, while there is one syllable between them in words with accent on the fourth syllable. Thus, there are differences in the proximity of tones. As a consequence of these differences, the LHL sequence of initial-accented words has to be realised within the first two syllables (or even within one syllable when the first syllable is long), while the LHHL sequence of the words with accent on the fourth syllable can be realised within four syllables. It may be possible to claim that, because of the difference in tightness, the alignment in initial-accented words is consistent and less variable compared to that in non-initial-accented words.

Another possible answer is based on the distance from the left-hand phrase boundary. The pitch accent, H*+L, in initial-accented words, for example, is adjacent to the left-hand phrase boundary, while it is one syllable away from the boundary in words with accent on the second syllable. The pitch accent is two and three syllables away from the phrase boundary in words with accent on the third syllable and on the fourth syllable, respectively, though there is a phrasal H- between them. As a result, the F0 rises from the boundary low to the pitch accent through different intervals depending on the tonal sequence, which may affect the alignment of the F0 peak for the pitch accent: the longer the time interval the F0 rise goes through, the earlier the F0 peak tends to be aligned.

Both these explanations, however, seem inadequate. It is widely reported that when two similar tones are too close to each other, they try to moderate the clashing situation in various ways such as pushing the neighbouring tone away (e.g. Silverman and Pierrehumbert 1990). Even though the phrasal H and the pitch accent H*+L are next to, or just one syllable away from, each other in words with accent on the third syllable and on the fourth syllable, the F0 peak is aligned earlier in words with accent on the third syllable and on the fourth syllable. That is, the accentual F0 peak was shifted leftward, rather than rightward. The account based on tightness seems contradictory in this respect. On the other hand, if the distance from the left-hand phrase boundary makes a difference, as suggested by the second explanation, there should be a gradual earlier shift of the F0 alignment in words with accent on the fourth syllable than in those with accent on the third syllable. However, as described in Chapter 5, there was no significant difference between them. Although it is very likely that the differences of tonal structures are involved in these alignment differences (or some other factor may come into play), it seems difficult to investigate any further with the data available in the current study.

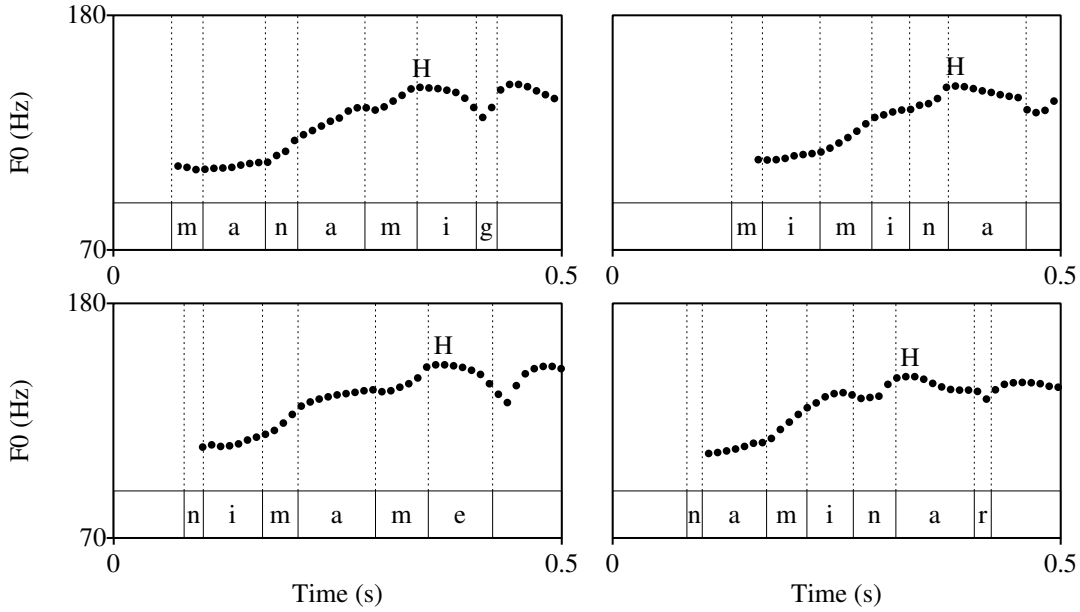


Figure 5.4: Examples of the F0 peak location for the phrase tone of an unaccented accentual phrase whose initial syllable is CV+CV and CVCV: /manami/ '(female name)', /miminari/ 'ringing in ears', /nimame/ 'boiled beans' and /naminari/ 'the roar of the waves'.

5.2.2 Unaccented accentual phrase

Visual inspection

Figure 5.4 shows examples of the F0 peak alignment for the phrasal H- in the CV+CV and CVCV sequences.[6] It can be seen that the F0 peak was aligned with the beginning of the third mora, which was observed in almost all the cases of the data. Figure 5.5 shows examples of CVN. The F0 peak was aligned in the vicinity of the beginning of the third mora. Figure 5.6 shows examples of CVR and CVV. Again, the F0 peak was aligned with the beginning of the vowel of the third mora. As in the data of CV+CV and CVCV, it was observed in most cases that the alignment of the F0 peak for the phrasal H- for CVN, CVR and CVV was fairly consistently aligned with the beginning of the vowel of the third mora. Thus, peak delay (i.e. ososagari) occurred in all the structures. Based on the patterns between the syllable/mora structures observed above, it seems reasonable to

[6]It is assumed that the phrasal H- is associated with the second mora in all the syllable/mora structures. See Section 2.2.1 in Chapter 2 for the full details of the material.

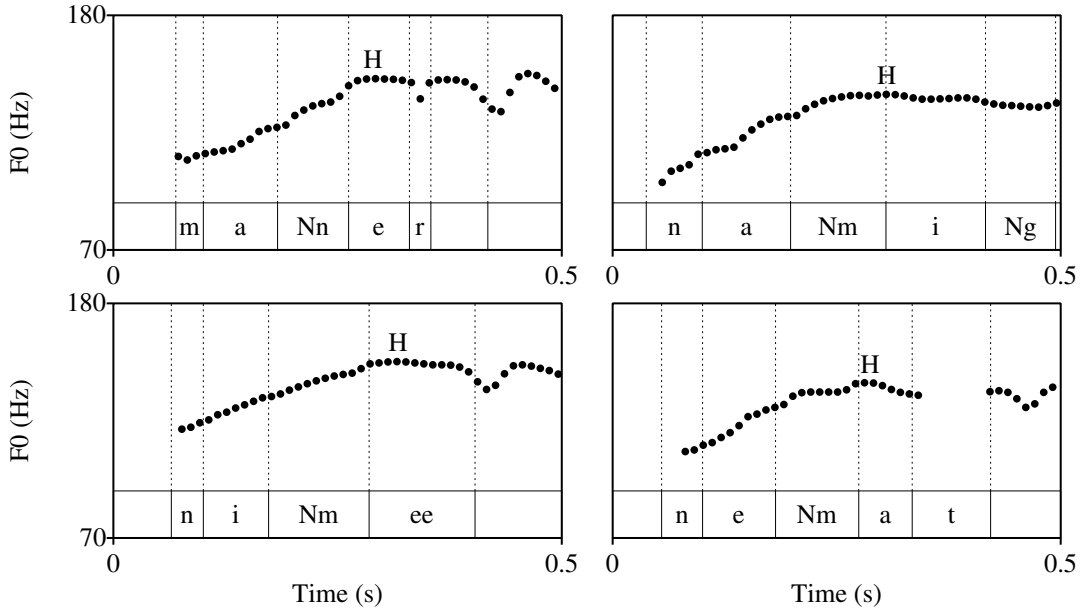


Figure 5.5: Examples of the F0 peak location for the phrase tone of an unaccented accentual phrase whose initial syllable is CVN: /maNneri/ ‘mannerism’, /naNmiN/ ‘refugee’, /nimeNsee/ ‘bilateral’ and /neNmatu/ ‘the end of the year’ (N is a moraic nasal).

use the beginning of the vowel of the following (third) mora as a segmental measurement point for the alignment of the F0 peak for the phrasal H-.[7]

Alignment of the F0 maximum for the phrasal H-

Figure 5.7 shows the alignment of H to the onset of the vowel of the third mora. The F0 peak was on average aligned 19 ms after the beginning of the vowel of the third mora in the CV+CV; 22 ms after the beginning of the vowel of the third mora for CVCV; 11 ms after the beginning of the vowel of the third mora for CVN; 2 ms after the beginning of the vowel of the third mora for CVR; and 19 ms after the beginning of the vowel of the third mora. The data were analyzed in a two-way (5×4) mixed design ANOVA, with items as the random factor, Structure (structures of the target syllable) as a within-items factor, and Speaker as a between-items factor. The ANOVA showed that there was no significant effect of Structure: $F(4, 96) = 1.663; p = 0.165$: no significant interaction

[7]There were quite a few cases where the onset consonant of the fourth mora was an obstruent, and it was often difficult to track the F0 movement around it. For example, it can be seen in /moomoku/ and /maemuki/ in Figure 5.6. In these adverse cases, the F0 was perturbed, and thus the F0 peak specification became unavoidably less reliable. At any rate, the best possible candidate was selected via a parabolic interpolation in these cases.

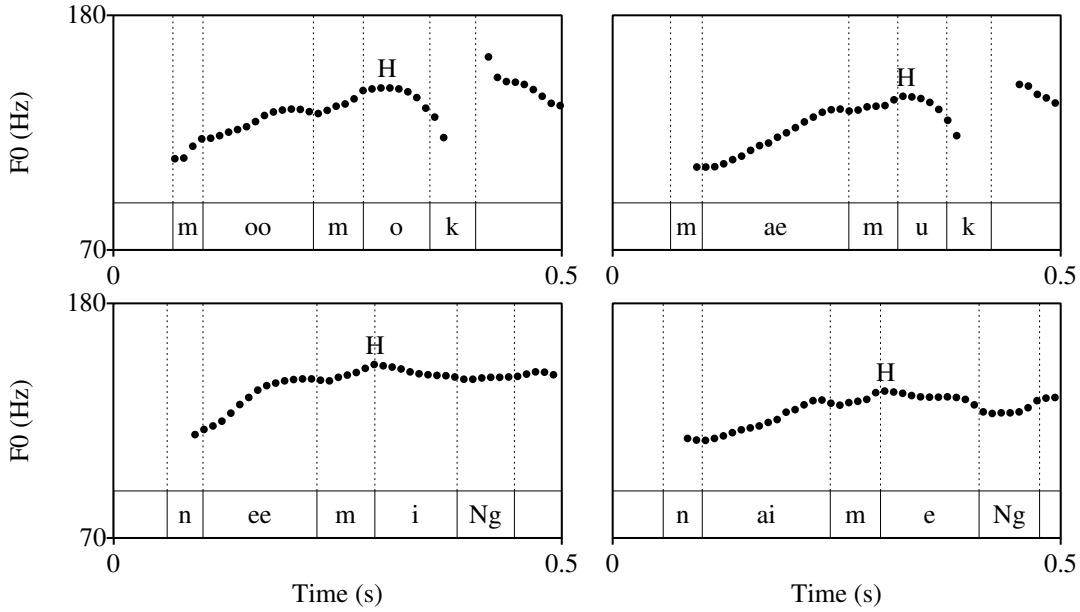


Figure 5.6: Examples of the F0 peak location for the phrase tone of an unaccented accentual phrase whose initial syllable is CVR and CVV: /moomoku/ ‘blind’, /maemuki/ ‘facing front’, /neemiNgu/ ‘naming’, and /naimeN/ ‘inside’.

between Structure and Speaker: $F(12, 96) = 1.598$; $p = 0.105$. There was no significant effect of Speaker: $F(3, 24) = 2.998$; $p = 0.51$.

The results showed that the F0 peak for the phrasal H- was aligned slightly after the beginning of the vowel of the third mora (i.e. the mora following the mora with which the phrasal H- is associated) in all the syllable/mora structures. Moreover, as discussed in (3.2) of Section 3.2.2 in Chapter 3, it can be regarded on Hayes’ terms that the F0 peak was anchored to the beginning of the second mora, and this is consistent with the alignment patterns found in the data of the previous chapters. To my knowledge, there is no quantitative study so far on the regularity of the F0 peak alignment for the phrasal H-. Some studies state that the F0 peak is delayed to the following mora or later, and that it is influenced by various factors (e.g. Venditti 2005). The current data clearly demonstrated fairly consistent alignment of the F0 peak for the phrasal H- with the beginning of the vowel of the following mora.

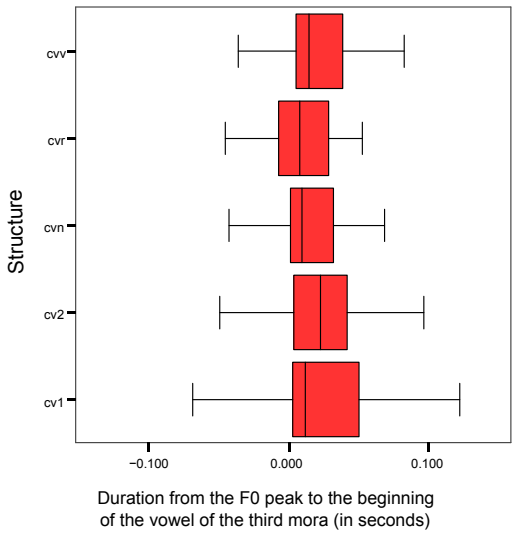


Figure 5.7: Boxplots of the alignment of H with the beginning of the vowel of the third mora (in seconds). The value zero in the graph corresponds to the beginning of the vowel of the third mora

5.3 Summary of the findings

On the data of the accented accentual phrase:

- The F0 peak was aligned around the end of the accented syllable in the word with accent on the second syllable, while it was aligned somewhere in the middle of the vowel of the accented syllable in the word with accent on the third or fourth syllables.
- There was a tendency for the accentual F0 peak to occur earlier relative to the accented syllable when the pitch accent was associated with a later syllable.
- There was a relatively large variation in the F0 peak alignment, compared to the data of initial-accented words.

On the data of the unaccented accentual phrase:

- The F0 peak for the phrasal H- was aligned slightly (about 15 ms) after the beginning of the vowel of the third mora of the accentual phrase in all the syllable/mora structures.

Chapter 6

Discussion and conclusion

I have been investigating tonal alignment in Tokyo Japanese. As stated in Chapter 1, the present study has two chief goals. One is to provide a thorough description of tonal alignment in Tokyo Japanese, including a well-known phenomenon, *ososagari*. Since there has been no large study on tonal alignment, particularly on the regularity of tonal alignment, in Tokyo Japanese so far, I have attempted to fill the gaps in our knowledge concerning the alignment of the F0 targets in Tokyo Japanese. The other goal is to contribute to the current understanding of tonal alignment, based on the empirical data of tonal alignment in Tokyo Japanese. As reviewed in Chapter 1, a large amount of evidence for regularities on tonal alignment in various languages have been provided recently, though there is still much disagreement on the characterisations and modelling of these alignment regularities. I have sought to provide further evidence to improve our understanding of tonal alignment. I discuss below to what extent these two goals were achieved.

6.1 Summary of the findings

The experiment described in Chapter 3 provides the data of the alignment of the F0 targets at the beginning of the initial-accented word. As discussed in Chapter 3, these data are most comparable to the data of the previous studies on tonal alignment in other languages and and most relevant to *ososagari* (a well-known tonal alignment phenomenon in Japanese). Examining the alignment of F0 targets in different syllable/mora structures of the accented syllable, I attempted to reveal how the F0 targets are aligned with the target segmental sequence, and whether there is any sort of alignment regularity. The overall results of the experiment indicate that both the F0 valley and peak are consistently

aligned with specific segmental landmarks, and the alignment of the F0 peak depends on the syllable/mora structure of the accented syllable. The F0 peak is aligned with the beginning of the vowel of the syllable following the accented syllable for CV+CV and CVCV; with the end of the first mora of the accented syllable for CVN ; and about 70 percent of the two-mora vowel of the accented syllable for CVR and CVV (see Section 2.2.1 of Chapter 2 for the use of the labels). These alignment patterns are seen across the speakers and regardless of the durational variation of the target segments.

In Chapter 4, I explored how the alignment patterns found in the experiment of Chapter 3 are influenced in different speaking modes; the speaking modes of interest were fast speech rate, raised voice, and local emphasis. Since some of the previous studies concerning tonal alignment in other languages show regularity of some aspects of the alignment of the F0 targets regardless of changes in a speaking mode, I assumed that some aspects of the alignment in Tokyo Japanese would also be unaffected by the changes in speaking modes. The results, on the whole, show that the orderly alignment behaviour seen in the data of Chapter 3 remain intact across different speaking modes. Firstly, there were basically no significant differences in alignment between the normal spoken data and the fast spoken data. Secondly, the F0 peak generally occurred slightly later in loud speech than in normal speech. Although the differences were very small, they were statistically significant in most conditions.[1] Despite the effect of overall pitch raising, it is also evident that the syllable/mora structure of the accented syllable plays an important role in the alignment, and that the alignment patterns are comparable to those in normal and fast speech. In effect, while the segmental durations are considerably influenced by the speech rate change and the overall pitch range is lifted up by raising the voice, the alignment is rather consistent irrespective of changes of these two speaking modes.

On the other hand, in speech with local emphasis, the F0 peak was aligned much earlier than in normal speech, and there were statistically significant differences between them. As reported in Chapter 4, because the speakers took different procedures when producing an utterance with local emphasis (e.g. putting a pause before and/or after the target word, or inserting an extra high tone around the end of the target word), more factors are involved in affecting the alignment and durational patterns. Among the factors involved, an extra high tone around the end of the target word plays an important role in the alignment by shifting the F0 peak leftward. However, it is also worth noting that, even in the data

[1] As shown in Chapter 4, overall pitch raising influenced overall F0 movements (which could be seen in the F0 trace as the less steep accentual F0 fall and, as a consequence, as the less acute F0 peak for the pitch accent). This seems to affect the automatic F0 peak specification in the form of the later F0 peak alignment.

of local emphasis, the alignment patterns depending on the syllable/mora structure of accented syllable can be detected, comparable to those seen in the other speaking modes.

In Chapter 5, I presented data on unaccented and non-initial-accented words, and compared the alignment patterns found in these data to those found in the data of initial accented words. It has been suggested that ososagari is less likely to occur with non-initial-accented words (Sugito 1982), and there is no accepted knowledge about the alignment with unaccented words. The results of the unaccented word demonstrated consistent alignment of the F0 peak with the beginning of the vowel of the syllable following the syllable with which the phrase H- is associated. It has been claimed that the F0 peak for the phrase H- may occur after the associated mora, but it is easily influenced by various factors such as information status and speech rate (e.g. Venditti 2005). The results of Chapter 5 confirmed ososagari—the F0 peak occurs after the associated mora—as previous studies claim. However, the results also demonstrated that the F0 peak for the phrase H- is consistently aligned with a specific segmental landmark. The F0 peak is actually aligned with the beginning of the vowel of the following syllable, like the alignment of the accentual F0 peak demonstrated in Chapters 3 and 4. The results therefore provide a more accurate description of the alignment of the F0 peak for the phrase H-.

As for the data of non-initial-accented words, I assumed, based on the results of Chapter 3, that the alignment of the accentual F0 peak would become similar to that of initial-accented words. The results of Chapter 5, however, showed that the F0 peak alignment of non-initial-accented words was not identical to that of initial-accented words presented in Chapter 3. The F0 peak was aligned not with the proposed segmental landmark, but somewhat earlier. It occurred earlier in words with accent on the third or fourth syllable than in words with accent on the second syllable, which, in turn, had the F0 peak earlier than initial-accented words. There was also much more variation of the alignment in the data of the non-initial-accented words than that of the initial-accented words. As discussed in Chapter 5, There are two explanations proposed for the alignment differences between the initial-accented word and the non-initial-accented word. One is based on overall 'tightness' between the tonal events due to the differences of prosodic structure; the other is based on the distance from the left-hand phrase boundary. However, it is also pointed out that the both explanations seem inadequate. Although it is very likely that the differences of tonal structures are involved in these alignment differences (or some other factor may come into play), it seems difficult to investigate any further with the data available in the current study.

In the remainder of this chapter, I will discuss issues raised from the results of the three experiments of the current study, and then point out several directions of future research.

6.2 Alignment under global changes

As summarised in Section 6.1 above, overall results of Chapter 4 show that the orderly alignment behaviours found in normal spoken data discussed in Chapter 3 remain unaffected regardless of speaking modes. Firstly, the alignment is basically not influenced under changes of speech rate. This is comparable to the findings from recent studies on other languages (Xu 1998; Ladd *et al.* 1999). Although Silverman and Pierrehumbert (1990) claim small effects of speech rate on alignment (i.e. the F0 peak is slightly later in fast speech than in normal speech) based on the result that it makes their model a slightly better fit in the multiple regression model, it can also be clearly seen that there is synchronisation between segments and F0 targets. Thus, based on the findings from this and other alignment studies, it seems that, while speech rate could have small effects on alignment, the alignment is rather consistent under speech rate changes. Secondly, as it is reported that both F0 peaks and valleys are affected by overall pitch raising, which can be thought that the whole pitch range is lifted up (e.g. Ladd and Terken 1995), it was supposed that, while the scaling of the F0 targets would be naturally affected (i.e. higher F0 values), the alignment was not, or was only little, affected by overall pitch raising. However, the F0 peak is aligned later in loud speech than in normal speech. One possible cause for this is that global F0 shapes were affected by the speakers' effort to speak as loud as possible. More specifically, the later peak was given rise to by a tendency in loud speech for the F0 be kept high after the initial rise at beginning of the accentual phrase. Thirdly, the F0 peak was aligned much earlier in speech with local emphasis than in normal speech. As described in 6.1 above, more factors are involved in speech with local emphasis, and their effects seem to override the normal segmental anchoring, which is comparable to clashing effects found in some of the previous studies (e.g. Silverman and Pierrehumbert 1990; Caspers and van Heuven 1993; Prieto *et al.* 1995).

The data discussed in Chapter 4 seem to emphasise the difficulty of explaining the alignment patterns found in fast speech rate by Xu and his colleagues' framework, as the normal spoken data above. One of the two situations in which Xu and Wang (2001) propose peak delay occurs is 'when the pitch target is [high] and surrounded by low pitch values and the duration of the host sufficiently short' (the other situation they propose is presumably irrelevant in Tokyo Japanese), and this seems to be the case for the target

tonal sequence, L H*+L, of the data in Chapter 4. Nonetheless, peak delay occurs regardless of the changes of segmental duration due to speech rate changes. It may still be possible to claim that peak delay occurs even in the normal spoken data because the duration of the host is not long enough, but I think this is a less plausible explanation, compared to segmental anchoring, which successfully generalises the alignment patterns found in the data of Chapters 3 and 4.

6.3 Alignment and syllable structure

As reviewed in Section 1.3, previous studies, particularly by Ladd and his colleagues, showed alignment regularities mainly with the data on the alignment of a bitonal pitch accent (such as L+H*). The pitch accent of interest in the data of Chapter 3 (and Chapter 4) is a bitonal pitch accent preceded by a low tone (L H*+L). Thus, it was to be expected that the valley and peak for the F0 rise (L and H* of H*+L) would be independently aligned with the segmental string. Nonetheless, what is comparable about the findings from the data in Chapter 3 is that the F0 peak for the pitch accent (H*+L) is consistently aligned with a specific segmental landmark depending on the syllable/mora structure of the accented syllable. Provided that an onset consonant is attached directly to a syllable node, these alignment patterns can be interpreted as a consequence of *segmental anchoring*: the F0 peak is anchored to the beginning of the second mora. This alignment regularity conforms to the findings from a series of work by Ladd and his colleagues. In addition, the data discussed in Chapter 3 also demonstrates phonological relevance to the alignment of the F0 targets as the Dutch data did in studies by Ladd *et al.* (2000) and Schepman *et al.* (2006), although the type of phonological factor is different—syllable/mora structure for Tokyo Japanese and phonological vowel length for Dutch.

As for *ososagari*, the data presented in Chapter 3 provide a more accurate description. It is generally stated that ososagari tends to occur in initial accented words whose second mora has a non-high vowel (Sugito 1982). However, as pointed out in Chapter 1, since it is based on observations of a very few samples, this characterisation is only partial and fragmentary. The results of Chapter 3 show that ososagari is not just a tendency of the phonetic realisation of a pitch accent. It is rather an orderly alignment behaviour, and, as noted above, it can be regarded as one of the consequences of segmental anchoring.

The alignment patterns found in the data in Chapter 3 provide evidence against the framework developed by Xu and his colleagues (e.g. Xu and Wang 2001). It is assumed in their

framework that a pitch target is synchronised with its 'host': it starts at the beginning of the host and ends at the end of the host. It is also assumed that a pitch target is realised through a process of continuous and asymptotic approximation during the duration of the host. Since it is normally considered that the mora is the tone-bearing unit (and the syllable is the accent-bearing unit) in Tokyo Japanese, the mora can be considered as the 'host' in Tokyo Japanese, and a pitch target (e.g. 'high') can be presumed to start at the beginning of the mora and end at the end of the mora. The results of Chapter 3, however, show that the 'high' does not end at the offset of the mora as Xu and his colleagues' model predicts; rather, it ends at different segmental points depending the syllable/mora structure of the accented syllable. As shown in Figure 3.10 in Chapter 3, while the F0 peak for the pitch accent in CVN and CVV can be regarded as ending at the end of the mora as suggested, the peak in CVCV ends at the beginning of the vowel of the following syllable. Even if we consider the syllable as the 'host' instead, the end of the pitch target does not synchronise with the end of the syllable; in fact, none ends at the offset of the syllable. In sum, the F0 peak ends neither at the offset of the mora nor the offset of the syllable.[2]

The results of Chapter 3 also have an implication for the syllable/mora structure in Tokyo Japanese. As reviewed in Chapter 1, there are various kinds of evidence that both the syllable and the mora play an important role in the phonology of Tokyo Japanese: the syllable in word formation and syllable weight; the mora in speech timing, phonological quantity, speech perception and so on. On the other hand, the internal structure of the syllable is still debatable. There are two important choices which make syllable-internal structures different from each other. One is whether the syllable has the rhyme as an independent constituent, that is, either whether the mora replaces the rhyme, or whether the rhyme is on a different metrical plane from that on which the mora is. The other is, if the mora is posited as a constituent in the syllable structure, whether the onset consonant is attached to the mora or directly to the syllable. Since it is almost indisputable as to the significance of the mora in Tokyo Japanese, and also since it seems that there is no evidence for the significance of the rhyme which has the peak and the coda as its constituents in Tokyo Japanese (i.e. the mora can completely replace the rhyme), the only relevant question seems to be whether the onset consonant is attathed to the mora, or directly to the syllable. As described in Section 6.1 above, the results of Chapter 3 show

[2]Furthermore, even if the accentual F0 fall is viewed as the implementation of the pitch target 'fall' (there are dynamic pitch targets, as well as static, proposed in Xu and Wang (2001)'s framework), their model fails to predict the alignment patterns found in Chapter 3, because the F0 fall for the pitch target 'fall' in CVCV begins at the beginning of the vowel of the following syllable, which is well after the offset of the accented syllable. As stated above, one of the implementation rules in their framework is that a pitch target is synchronised with the 'host'.

that there are three-way alignment patterns of the accentual F0 peak: it is aligned with the beginning of the vowel of the syllable following the accented syllable for CV+CV and CVCV; with the end of the first mora (i.e. the end of the vowel) of the accented syllable for CVN; and in the middle of the two-mora vowel of the accented syllable for CVR and CVV. These alignment behaviours can be generalised, given that the onset consonant is directly attached to the syllable. On the other hand, it is difficult to generalise the alignment patterns across the different structures of the accented syllable, given the onset consonant is attched to the mora. Since the internal structure of the syllable provided leads to a better explanation for the alignment patterns due the different syllable/mora structures found in Chapter 3, it seems possible to claim that the results indirectly supports the validity of the internal structure of the syllable in which onset consonants are attached directly to a syllable node, as proposed in moraic phonology (e.g. Hayes 1989).

As reviewed in Chapter 1, there are two competing proposals for accounting for alignment patterns across languages. One is based on the *phonetic continuum of alignment* (Atterer and Ladd 2004). The other is *phonological anchoring* (Prieto *et al.* forthcoming). Considering the results of Chapters 3 and 4, there are implications for these proposals. Phonological anchoring is originally proposed in order to explain language-specific contrasts of alignment in Romance languages, and can also be applicable to the description of alignment regularities in other languages. Although it makes no difference which metrical edge H* is secondarily associated with because Tokyo Japanese has only one pitch accent type, H*+L, and no contrast of alignment, H* of H*+L can be regarded as being secondarily associated with the left edge of the second mora, based on the results of Chapter 3. This actually ends up in two different autosegmental representations, however. One is where H* is associated with the left edge of the mora in the syllable with which the pitch accent is primarily associated; the other is where H* is associated with the left edge of the mora in the syllable after the accented syllable. While it is not clear whether this is allowed in Prieto *et al.* (forthcoming), there are two different representations provided for one type of pitch accent. On the other hand, there seems to be no such incompatibility with the proposal by Atterer and Ladd (2004). It can be considered that Tokyo Japanese chooses a distinct place along the phonetic continuum of alignment in a language-specific way. Nevertheless, there is still an open question as to why Tokyo Japanese chooses this specific location of prosodic structure.

6.4 Timing control and synchronisation

One of the related findings from the data of Chapter 3 is that the vowel duration of the accented syllable varies considerably depending on the syllable/mora structure of the accented syllable. The vowel duration of the accented syllable of .**CV**.CV. is about half of that of .**CVN**.CV., which, in turn, is about two thirds of that of .**CVV**.CV. (the accented syllable is shown in bold type). Note that, even though the first 'CV's in .**CV**.CV. and .**CVN**.CV. can be considered to have the same length phonologically, the duration of the first vowel of .**CVN**.CV. is much longer than that of .**CV**.CV. Further durational analyses reveal that 'V.C' in .**CV**.CV., 'V' in .**CVN**.CV., 'NC' in .**CVN**.CV., the first 'V' in .**CVV**.CV. (which is the first mora of the two-mora vowel), and the 'V.C' in .**CVV**.CV. (which is the second mora of the two-mora vowel and the onset consonant of the following syllable) are all similar in duration. Based on these results, I propose that 'V.C' (or 'V', if there is no onset consonant in the following mora) and a moraic consonant work as the timing unit in Tokyo Japanese, rather than '.CV.' (or 'V', when it is the second mora of a two-mora vowel) and a moraic consonant, which has normally been regarded as the timing unit.

Although it has been reported that vowels in a syllable with a moraic consonant in the coda position are longer than those in a light syllable (e.g. Homma 1981), the findings described above, together with those from the data of tonal alignment, have an interesting implication. The results of Chapter 3 show that the F0 valley for the boundary L is aligned with the beginning of the accented syllable (which is the left edge of the target accentual phrase), while the F0 peak for the pitch accent is aligned with the beginning of the second mora. Since the segmental members between these segmental landmarks vary depending on the syllable/mora structure of the accented syllable, the time between the landmarks for the F0 rise may not be long enough if the number of the segments is too small or their intrinsic duration is too short. I suspect that this is the case for the rise of the .CVN.CV., because the F0 peak is aligned with the beginning of the moraic nasal and there are only an onset consonant and a vowel between the segmental landmarks. As a consequence, the vowel is lengthened in order to give it enough time to reach the high target. Thus, the longer vowel duration of the syllable with a moraic nasal at the end results from the mutual temporal coordination between segments and tones.

6.5 Limitations of the current study and future directions

Finally, I suggest several directions for future research. First, as discussed above, factors producing the alignment differences between initial-accented words and non-initial words have to be clarified. Differences in tonal structure are likely to be involved, though the data of the current study thorough enough, particularly in terms of the syllable/mora structures of the accented syllables in non-initial-accented words. Further research with appropriate material will give us a better picture of alignment regularity in Tokyo Japanese. Secondly, the cause of the longer vowel duration of the syllable with a moraic consonant in the coda position has to be fully investigated. I suggest that this longer vowel duration is one of the consequences of the mutual synchronisation between segments and tones. If this interaction is well established, it can be considered as strong evidence for how significant segmental anchoring is in the overall timing control in Japanese speech. Thirdly, the segmental duration with the domain proposed in the current study has to be explored with appropriate data. A vowel and the onset consonant of the following mora ('V.C'), rather than the onset consonant in the same mora ('CV'), seem to work as the timing unit in Japanese speech. Although the findings from the data of Chapter 3 are still preliminary, this is a novel analysis of mora-timing, a longstanding issue in Japanese phonetics and phonology, and further research may give us solid phonetic data about what 'mora-timing' actually is. Fourthly, the significance of the alignment regularity found in the current study in speech perception needs to be investigated. In the present work, tonal alignment in Tokyo Japanese has solely been described in terms of speech production. A perceptual study based on the language-specific alignment patterns found in the current study is awaited. Finally, the effects of contextual factors on tonal alignment have to be thoroughly explored. As demonstrated with data in Chapter 4, different contextual factors affect tonal alignment in complicated ways, so it is necessary to study those, as well as alignment regularity, in order to achieve a deeper understanding of temporal coordination between segments and tones.

Appendix A

Tables of the data in Chapter 3

Speaker	CV+CV		CVCV		CVN		CVR		CVV	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AK	181.5	22.6	181.0	30.2	164.5	31.0	153.7	23.9	165.5	24.3
CE	147.6	18.7	156.9	29.3	168.8	21.9	151.7	26.5	172.2	22.6
FS	161.2	30.5	173.0	31.4	175.8	24.1	159.4	22.5	163.7	24.4
NI	174.3	29.0	176.4	30.9	163.4	37.0	162.1	39.5	171.0	39.9
RM	121.8	16.1	131.8	29.6	116.3	20.1	121.7	34.2	134.5	26.3
ST	150.7	25.9	168.0	23.2	165.8	20.8	163.5	21.8	160.9	21.9
TN	130.3	22.5	146.2	28.5	140.4	16.3	124.9	20.3	143.4	17.1
Total	152.8	30.7	162.6	31.8	156.5	31.0	148.1	31.2	159.1	27.5

Table A.1: Alignment of H relative to C0 in ms.

Speaker	CV+CV		CVCV		CVN		CVR		CVV	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
TN	30.9	30.2	45.7	25.6	−10.4	15.9	−65.0	19.0	−55.8	30.2
ST	52.8	15.4	58.2	15.2	−4.8	16.0	−44.3	16.9	−42.7	25.3
CE	58.2	7.2	64.8	16.4	8.0	15.6	−62.1	19.7	−41.2	20.5
RM	11.1	34.0	22.8	37.8	−28.3	12.3	−90.5	14.0	−88.2	34.3
FS	56.8	31.9	64.7	28.7	19.4	17.0	−38.9	13.2	−46.0	17.4
AK	45.9	39.3	64.1	12.1	8.6	20.4	−44.9	17.9	−43.6	18.0
NI	67.0	17.3	71.0	25.1	6.3	30.9	−25.9	11.7	−31.6	15.0
Overall	46.9	30.0	56.9	25.6	−0.2	22.3	−52.7	24.1	−48.2	27.1

Table A.2: Alignment of H relative to C1 in ms.

Speaker	CV+CV		CVCV		CVN		CVR		CVV	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
TN	−17.0	30.1	−10.9	27.2	−108.3	23.5	−104.3	21.5	−107.0	26.9
ST	−2.2	15.0	−6.1	19.0	−94.3	17.0	−84.3	24.1	−94.2	37.0
CE	10.3	7.6	11.8	7.4	−92.2	23.1	−87.9	16.4	−87.9	26.1
RM	−43.0	37.0	−32.2	38.9	−151.3	14.3	−143.3	18.3	−122.9	43.5
FS	14.9	32.9	12.6	32.0	−73.1	14.4	−72.7	14.4	−98.5	15.9
AK	8.1	13.3	−1.6	9.1	−99.3	19.4	−87.7	18.1	−90.8	30.5
NI	−13.0	75.0	17.7	25.0	−86.1	37.4	−71.7	22.5	−72.2	27.6
Overall	−5.1	37.7	−0.9	26.4	−100.2	29.7	−92.2	27.9	−94.9	32.2

Table A.3: Alignment of H relative to V1 in ms.

Speaker	CV+CV		CVCV		CVN		CVR		CVV				
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	df	F	P
AK	296	19	302	11	304	20	292	9	296	6	4,36	1.220	0.319
CE	307	13	300	13	295	12	292	13	299	17	4,36	1.368	0.264
FS	164	4	163	4	163	4	162	4	160	4	4,20	3.648	0.022
NI	272	8	267	7	268	7	267	7	268	6	4,28	2.486	0.066
RM	138	4	136	8	144	7	139	5	138	5	4,20	1.893	0.151
ST	259	10	257	10	263	9	254	9	259	11	4,36	0.949	0.447
TN	156	6	155	3	154	5	156	5	155	6	4,36	0.300	0.876
Overall	230	68	234	66	227	66	223	64	226	66			

Table A.4: Mean F0 peak values in Hz.

Structure	c0tov0	v0toc1	c1tov1	Total
CVV	4.8	15.8	4.5	25.1
CVR	5.2	15.2	3.9	24.2
CVN	5.1	10.6	10.0	25.7
CVCV	4.4	5.6	5.9	16.0
CV+CV	4.3	6.1	5.4	15.8

Table A.5: Mean duration of the segments of the target sequences in ms.

Structure	Mean	SD
CVV	0.5	0.8
CVR	−0.3	0.9
CVN	1.8	0.3
CVCV	−0.5	0.7
CV+CV	−0.1	0.7

Table A.6: Alignment of L relative to C0 in ms.

Appendix B

Tables of the data in Chapter 4

Speaker	ak			ce			fs			ni		
Speaking Mode	N	RV	LE	N	RV	LE	N	RV	LE	N	RV	LE
Mean	298.4	416.9	325.9	298.5	388.3	289.3	162.3	208.7	188.2	268.3	309.0	228.3
SD	14.4	14.0	13.9	14.3	17.7	13.0	4.1	7.0	7.4	7.1	8.3	11.7

Speaker	rm			st			tn		
Speaking Mode	N	RV	LE	N	RV	LE	N	RV	LE
Mean	139.4	189.3	159.9	258.3	359.4	303.5	154.9	246.4	175.1
SD	6.0	9.4	7.8	10.1	29.1	10.8	4.7	5.6	8.0

Table B.1: Mean F0 peak values in Hz between Normal (N), Raised Voice (RV) and Local Emphasis (LE). ‘fs’, ‘rm’ and ‘tn’ are male speakers, and the rest are female speakers.

Speaker	Normal		Raised		Fast		Local Emphasis	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ak	0.3	8.0	16.3	20.4	1.7	7.6	−0.3	9.1
ce	10.0	7.3	49.3	21.9	23.0	14.0	−25.8	27.4
fs	8.9	29.8	24.6	21.6	4.8	18.0	−13.1	34.9
ni	14.8	22.8	19.8	12.7	21.8	12.1	−41.9	38.8
rm	−29.1	21.9	7.9	9.9	−11.8	15.0	−11.2	11.4
st	−6.7	16.7	39.6	27.3	4.2	14.5	−33.2	29.9
tn	−6.3	20.4	9.2	13.5	−6.5	27.7	4.4	16.1
Total	−0.6	23.3	22.5	22.6	5.1	20.1	−17.1	30.8

Table B.2: Alignment of H relative to V1 in ms for CV+CV and CVCV across the different speaking modes.

Speaker	Normal		Raised Voice		Fast		Local Emphasis	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ak	5.1	21.3	9.4	18.8	−0.1	13.7	−7.1	10.8
ce	8.0	15.6	−0.3	15.5	19.0	20.1	−34.4	20.7
fs	17.9	13.1	41.4	16.2	18.4	13.2	−7.7	12.4
ni	6.8	29.0	−6.7	5.5	−10.5	3.9	−28.0	9.0
rm	−27.3	11.1	−11.7	11.5	−17.0	10.4	−17.5	10.4
st	−5.0	16.9	3.2	13.0	16.3	14.1	−16.1	9.6
tn	−10.4	15.9	10.1	44.2	−12.0	5.3	−9.2	4.8
Total	−0.6	22.3	5.9	25.9	1.9	18.9	−17.8	15.3

Table B.3: Alignment of H relative to C1 in ms for CVN across the different speaking modes.

Speaker	Normal		Raised Voice		Fast		Local Emphasis	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ak	−45.2	17.0	−18.8	24.2	−26.3	10.3	−62.8	31.4
ce	−49.5	20.8	−17.5	11.4	−21.7	12.1	−97.8	31.0
fs	−51.7	26.3	−37.7	29.8	−32.9	26.7	−64.1	20.9
ni	−27.0	12.1	−17.4	18.6	−18.8	5.3	−65.3	19.7
rm	−92.7	22.6	−43.5	41.4	−43.8	25.1	−79.3	31.0
st	−43.5	21.0	−20.5	31.4	−13.7	19.4	−101.5	30.6
tn	−58.5	26.6	−52.5	32.2	−45.4	21.8	−77.0	19.8
Total	−52.1	27.8	−29.1	30.6	−28.2	21.5	−79.5	30.5

Table B.4: Alignment of H relative to C1 in ms for CVR and CVV across the different speaking modes.

Appendix C

Tables of the data in Chapter 5

Second		Third		Forth	
Mean	SD	Mean	SD	Mean	SD
23.7	32.3	−32.5	46.2	−43.7	28.5

Table C.1: Alignment of H with the end of the accented syllable in ms.

Second		Third		Forth	
Mean	SD	Mean	SD	Mean	SD
86.5	34.9	62.5	42.6	35.1	33.1

Table C.2: Alignment of H with the beginning of the vowel of the accented syllable in ms.

CV+CV		CVCV		CVN		CVR		CVV	
Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
19.1	9.4	17.9	7.5	10.5	6.8	1.8	9.2	19.0	5.4

Table C.3: Alignment of H with the beginning of the vowel of the third mora in ms.

References

Arvaniti, Amalia and D. Robert Ladd. 1995. Tonal alignment and the representation of accentual targets. In Elenius, Kjell and Peter Branderud, eds., *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 4, pp. 220–3. Stockholm.

Arvaniti, Amalia, D. Robert Ladd and Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3–25.

Arvaniti, Amalia, D. Robert Ladd and Ineke Mennen. 2000. What is a starred tone? In Broe, Michael B. and Janet B. Pierrehumbert, eds., *Language Acquisition and the Lexicon*, Papers in Laboratory Phonology V, pp. 119–131. Cambridge: Cambridge University Press.

Atterer, Michaela and D. Robert Ladd. 2004. On the phonetics and phonology of "segmental anchoring" of *F0*: evidence from German. *Journal of Phonetics* 32: 177–197.

Beckman, Mary E. and Janet B. Pierrehumbert. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3: 255–309.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, vol. 17, pp. 97–110. Amsterdam, Netherlands.

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Caspers, Johanneke and Vincent J. van Heuven. 1993. Effects of time pressure on the phonetic realisation of the accent-lending pitch rise and fall. *Phonetica* 50: 161–71.

D'Imperio, Mariapaola. 2000. The role of perception in defining tonal targets and their alignment. PhD thesis, Ohio State University.

D'Imperio, Mariopaola. 2001. Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33(4): 339–356.

Face, Tim. 2001. Focus and early peak alignment in Spanish intonation. *Probus* 13: 223–246.

Fujisaki, Hiroya. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In MacNeilage, Peter F., ed., *The Production of Speech*, pp. 39–55. New York: Springer-Verlag.

Gili Fivela, Barbara. 2002. Tonal alignment in two Pisa Italian peak accents. In Bel, B. and I. Marlien, eds., *Proceedings of the Speech Prosody 2002 Conference*, pp. 339–342. Aix-en-Provence.

Goldsmith, John A. 1975. Autosegmental phonology. PhD thesis, Massachusetts Institute of Technology.

Grice, Martine, D. Robert Ladd and Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17: 143–185.

Gussenhoven, Carlos. 2000. The boundary tones are coming: on the nonperipheral realization of boundary tones. In Broe, Michael B. and Janet B. Pierrehumbert, eds., *Language Acquisition and the Lexicon*, Papers in Laboratory Phonology V, pp. 132–151. Cambridge: Cambridge University Press.

Haraguchi, Shosuke. 1977. *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha.

Hasegawa, Yoko and Kazue Hata. 1992. Fundamental frequency as an acoustic cue to accent perception. *Language and Speech* 35: 87–98.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20: 253–306.

Homma, Yayoi. 1981. Durational relationships between Japanese stops and vowels. *Journal of Phonetics* 9: 273–81.

House, Jill, Jana Dankovičová and Mark Huckvale. 1999. Intonation modelling in ProSynth: an integrated prosodic approach to speech synthesis. In Ohala, John J., Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey, eds., *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 2343–2346. San Francisco.

Kagomiya, Takayuki. 1998. Kutoo no ehuzero joosyoo to akusento kaku ni yoru ehuzero kakoo to no kankei [The relationship between the phrase-initial F0 rise and the F0 fall of pitch accent]. In *Proceedings of the 13th general meeting of the Phonetic Society of Japan*, pp. 97–102.

Knight, Rachael-Anne. 2002. The influence of pitch span on intonational plateaux. In Bel, Bernard and Isabelle Marlien, eds., *Proceedings of Speech Prosody 2002*, pp. 439–442. Aix-en-Provence, France.

Kubozono, Haruo. 1999. Mora and syllable. In Tsujimura, Natsuko, ed., *The Handbook of Japanese Linguistics*, pp. 31–61. Oxford: Blackwell Publishers.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge, England: Cambridge University Press.

Ladd, D. Robert. 2003. Phonological conditioning of F0 target alignment. In Solé, Maria-Josep, Daniel Recasens and Joaquín Romero, eds., *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 249–252. Barcelona, Spain: Causal Productions.

Ladd, D. Robert, D. Faulkner, H. Faulkner and Astrid Schepman. 1999. Constant segmental anchoring of *F*0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106: 1543–54.

Ladd, D. Robert, Ineke Mennen and Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107: 2685–96.

Ladd, D. Robert and Jacques Terken. 1995. Modeling inter- and intra-speaker pitch range variation. In Elenius, Kjell and Peter Branderud, eds., *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 2, pp. 386–389. Stockholm, Sweden.

McCawley, James D. 1968. *The Phonological Component of a Grammar of Japanese*. The Hague, The Netherlands: Mouton.

Neustupný, Jiri Vaclav. 1966. Is the Japanese accent a pitch accent? *Onsei Gakkai Kaihoo* 121: 1–7. (In Japanese).

NHK, ed. 1998. *The Japanese Language Pronunciation and Accent Dictionary*. Tokyo: Japan Broadcast Publishing Co., 2nd edn. (In Japanese).

Nooteboom, Sieb G. and I. H. Slis. 1972. The phonetic feature of vowel length in Dutch. *Language and Speech* 15: 301–316.

Pierrehumbert, Janet B. 1980. The phonology and phonetics of English intonation. PhD thesis, Massachusetts Institute of Technology.

Pierrehumbert, Janet B. and Mary E. Beckman. 1988. *Japanese Tone Structure*. Cambridge, Massachussetts: MIT Press.

Poser, William J. 1984. The phonetics and phonology of tone and intonation in Japanese. PhD thesis, Massachusetts Institute of Technology.

Prieto, Pilar, Mariapaola D'Imperio and Barbara Gili Fivela. forthcoming. Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech* .

Prieto, Pilar, Jan P. H. van Santen and Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23: 429–51.

Schepman, Astrid, Robin Lickley and D. Robert Ladd. 2006. Effects of vowel length and "right context" on the alignment of dutch nuclear accents. *Journal of Phonetics* 34(1): 1–28.

Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge Language Surveys. Cambridge: Cambridge University Press.

Shinya, Takahito and Miyuki Takasawa. 1999. Nihongo no hatuwa matu ni okeru ehuzero kakoo no arainmento ni tuite [On the alignment of the utterance-final F0 fall in Japanese]. In *Proceedings of the 14th general meeting of the Phonetic Society of Japan*.

Silverman, Kim and Janet B. Pierrehumbert. 1990. The timing of prenuclear high accents in English. In Kingston, John and Mary E. Beckman, eds., *Between the Grammar and Physics of Speech*, Papers in Laboratory Phonology I, pp. 71–106. Cambridge: Cambridge University Press.

Sugito, Miyoko. 1982. *Nihongo Akusento no Kenkyuu [Studies on Japanese Accent]*. Tokyo: Sanseido.

Venditti, Jennifer J. 1997. Japanese ToBI labelling guidelines. *Ohio State University Working Papers in Linguistics* 50: 127–162.

Venditti, Jennifer J. 2005. The J_ToBI model of Japanese intonation. In Jun, Sun-Ah, ed., *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press. Collection of papers from the ICPhS 1999 satellite workshop on 'Intonation: Models and ToBI Labeling'. San Francisco, California.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61–83.

Xu, Yi. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *Journal of Phonetics* 27: 55–105.

Xu, Yi. 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58: 26–52.

Xu, Yi and Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of Acoustical Society of America* 111(3): 1399–1413.

Xu, Yi and Q. Emily Wang. 2001. Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication* 33: 319–337.